Search for the Higgs Boson in the process H $\rightarrow$ ZZ $\rightarrow$ $ll\nu\nu$ produced via Vector-Boson Fusion with the ATLAS detector

Clive Andrew Edwards

Department of Physics

Royal Holloway, University of London

A thesis submitted to the University of London for the degree of Doctor of Philosophy

March 2012

Declaration

I declare that the work presented in this thesis is my own. Where information has been derived from other sources I confirm this has been indicated in the document.

Clive Andrew Edwards

Abstract

The search potential of a Standard Model Higgs boson in the Vector Boson Fusion production mechanism with Higgs boson decaying to two leptons and two neutrinos via decay to two Z bosons with the ATLAS detector is investigated. The ATLAS detector is a general purpose detector in operation at CERN measuring proton-proton collisions produced by the Large Hadron Collider. This channel has been shown to have high sensitivity at large Higgs mass, where large amounts of missing energy in the signal provide good discrimination over expected backgrounds. This work takes a first look at whether the sensitivity of this channel may be improved using the remnants of the vector boson fusion process to provide extra discrimination, particularly at lower mass where sensitivity of the main analysis is reduced because of lower missing energy.

Simulated data samples at centre of mass energy 7 TeV are used to derive signal significances over the mass range between 200-600 GeV/c^2. Because of varying signal properties with mass, a low and a high mass event selection were developed and optimized. A comparison between simulated and real data (collected in 2010) is made of variables used in the analysis and the effect of pileup levels corresponding to those in the 2010 data is investigated. Possible methods to estimate some of the main backgrounds to this search are described and discussed. The impact of important theoretical and detector related systematics are taken into account. Final results are presented in the form of 95 % Confidence Level exclusion limits on the signal cross section relative to the SM prediction as a function of Higgs boson mass, based on an integrated luminosity of $33.4\,\mathrm{pb}^{-1}$ of data collected during 2010.

Acknowledgements

I would like to thank my supervisor Pedro Teixeira-Dias for giving me the opportunity to work towards a PhD and his help and support towards achieving it. I gratefully acknowledge the financial support provided by the Science and Technology Facilities Council (STFC) during this degree and thank Royal Holloway, University of London for providing funding for several opportunities.

I would also like to thank all of the students, postdocs and academics in the Royal Holloway physics department for their help in solving many of the problems I have encountered. Special thanks to Ricardo for his guidance and Rudi for technical discussions while I was at CERN. Thankyou to the ATLAS Collaboration, who produced the Monte Carlo samples used in this thesis.

I am eternally grateful to my family for their love and tireless support and to Marine for all her sacrifices and for making me happy. Without you this thesis could not have been realised.

Contents

Preface

The goal of this study is to search for a Standard Model Higgs boson in the $H \to ZZ \to ll\nu\nu$ decay channel using the Vector Boson Fusion production mechanism at the Large Hadron Collider with the ATLAS experiment. This thesis is organised in the following way:

- Chapter 1 presents a summary of the Standard Model of particle physics and the motivation for a Higgs boson.
- Chapter 2 gives a short description of the LHC and then moves on to detail the ATLAS experiment, one of the tools it is hoped will provide experimental evidence for a Higgs type boson.
- Chapter 3 describes the signal and background processes relevant to this analysis. The method of Monte Carlo simulation is summarised and the details of the Monte Carlo samples used in this study are then outlined. Finally an overview of the data sample analysed is given.
- Chapter 4 describes the reconstruction of physics objects within ATLAS. The selection criteria for the objects used in the analysis are outlined and where relevant, choices made are motivated. The performance of the reconstruction of each of the physics object relevant to this study is investigated.
- Chapter 5 describes the analysis strategy to search for the signal. A baseline set of one dimensional cuts are obtained and the results presented.
- Chapter 6 describes the optimization procedures which were used to optimize the signal significance. Final results are compared to the baseline selection as a function of Higgs mass.
- Chapter 7 describes the effect of several systematic uncertainties on the analysis. The results of data-driven methods to estimate the main backgrounds identified in this search are summarised. An estimate of the total systematic uncertainty on the selections developed is given. Final results are presented in the form of 95 % Confidence Level exclusion limits on the signal cross section relative to the SM prediction as a function of Higgs boson mass.

Chapter 1

The Standard Model of Particle Physics

This chapter summarises the theoretical aspects relevant to this study. It is composed of two parts. First, our understanding of particle physics is reviewed in the context of the Standard Model and then, the Standard Model Higgs boson, the subject of this thesis is discussed.

1.1 Elementary particles and their interactions

Our current understanding of the physical world in terms of fundamental matter particles and their interactions is largely based on the Standard Model (SM) of particle physics. It describes all known particles and three of the four known fundamental interactions (the electromagnetic, weak and strong interactions). Within the SM the particles are classified by their spin as either

- half-integer spin particles called fermions obeying Fermi-Dirac statistics. These form the matter particles.
- integer spin particles called bosons. These particles obey Bose-Einstein statistics and their exchange between the fermions describes the fundamental interactions.

1.1.1 The Fermions

The fermions are categorised into two types, the quarks and the leptons. The quarks are given a baryon number B=1/3. The leptons are assigned a lepton number L=1 and do not interact strongly. Each type of fermion consists of three families or generations, each of which consists of two distinct particles (often referred to as a

Leptons				Quarks		
Generation	Particle	Mass	Charge	Particle	Mass	Charge
1	ν_e	$< 2\times10^{-9}$	0	u	0.0015-0.003	$+\frac{2}{3}$
1	e	5×10^{-4}	-1	d	0.0035-0.006	$-\frac{1}{3}$
2	ν_μ	$< 1.9\times10^{-9}$	0	c	1.27	$+\frac{2}{3}$
2	μ	0.105	-1	s	0.105	$-\frac{1}{3}$
3	ν_τ	$< 1.85\times10^{-5}$	0	t	172.9	$+\frac{2}{3}$
3	τ	1.78	-1	b	4.2	$-\frac{1}{3}$

Table 1.1: The fermions and some of their basic properties. Masses are given in GeV/c^2 and the electric charge in units of the electron's charge, $|e|$ [1]

doublet of particles). Table 1.1 lists the quarks and leptons and some of their basic properties [1].

The first generation of quarks consists of the up (u) quark with $+\frac{2}{3}$ electric charge (in units of electron charge, $|e|$) and the down (d) quark with -$\frac{1}{3}$ electric charge. The other generations consist of a u-type and a d-type quark but are successively heavier than the first generation. The second generation consists of the strange (s) and charm (c) quarks while the third generation consists of the bottom (b) and top (t) quarks. Quarks carry colour charge and as such each comes in three distinct colour states (red, green or blue). Each doublet of leptons is composed of an electrically charged lepton and its corresponding neutral neutrino. As with quarks, the mass of the charged leptons in the doublet increases with generation. The first generation consists of the electron (e) and its neutrino (ν_e), the second the muon (μ) and its neutrino (ν_μ) and the third the tau (τ) and its neutrino (ν_τ). Each quark and lepton have a corresponding anti-particle, denoted with a bar. Anti-particles have opposite electric charge to the corresponding particle but the same mass. In nature quarks are only found within composite hadrons, composed of either three quarks making a baryon or in quark anti-quark states called mesons.

1.1.2 The Bosons

Interactions between the fermions are mediated by the absorption and emission of integer spin particles called bosons. This gives rise to four fundamental forces, summarised in Table 1.2. The electromagnetic force makes the electron bind to nuclei and more generally, molecule formation underpinning Chemistry. It is mediated by the photon (γ). The strong force is responsible for holding nuclei together and is mediated by eight massless gluons (g). The weak force explains β decay and is mediated by exchange of $W^{+/-}$ and Z bosons. Gravity is responsible for galactic formation. It is the weakest of all the forces and is negligible at the energy scales considered in particle physics. In principle, gravity can be described as being mediated by the exchange of a boson called the Graviton.

Force	Mediator	Spin	Charge	Mass (GeV)	Relative Strength	Decays (%)
Electromagnetic	γ	1	0	0	10^{-2}	-
Strong	Gluons	1	0	0	1	-
Weak	W	1	± 1	80.4	10^{-9}	$\rightarrow l\nu$ (32.4) $\rightarrow$ hadrons (67.6)
Weak	Z	1	0	91.2	10^{-9}	$\rightarrow ll$ (10.1) $\rightarrow \nu\nu$ (20.0) $\rightarrow$ hadrons (69.9)
Gravity	Graviton	2	0	0	10^{-38}	-

Table 1.2: The mediators of the four fundamental fermion interactions and some of their basic properties. Relative strength corresponds to the typical strength of the forces compared to the strong force between two protons separated by $\approx 15\,\text{fm}$ [2]. Here l refers to lepton (e, μ and τ).

1.2 The Standard Model

The SM is a theoretical framework of quantum field theory [3] in which the elementary particles are the quanta of the underlying fields and the interactions are a consequence of the principle of local gauge invariance. As yet attempts to incorporate gravity using this approach have failed. The time-line of the SM becoming a unified theory of the forces that it describes started with the development of the quantum field theory of electromagnetic interactions, called Quantum Electrodynamics. Subsequently in the 1960's a unified electroweak theory was developed unifying the electromagnetic and weak interactions. Finally the electroweak theory was unified with the theory of the strong interactions (Quantum Chromodynamics) giving what is understood as the SM today.

1.2.1 Quantum Electrodynamics (QED)

All quantum electromagnetic interactions consist of the interaction of charged fermions with the quantum of the electromagnetic field, the photon. The most basic form of such an interaction, is shown in Figure 1.1. It shows the interaction of a charged fermion f with a photon γ. As with all interactions, the strength is characterized by a coupling constant associated to each vertex. The electromagnetic force couples to electric charge and so this defines the strength of electromagnetic interactions. This vertex corresponds to the basic building block from which all QED processes can be represented. Complete QED processes represented in this way are called Feynman diagrams. Feynman diagrams with the smallest number of vertices for a given process to occur are referred to as tree-level or leading-order whereas diagrams with a higher number of vertices are called higher order diagrams. A detailed picture of any QED process can be obtained by summing over all possible internal states and this corresponds to summing over all Feynman diagrams of all orders.

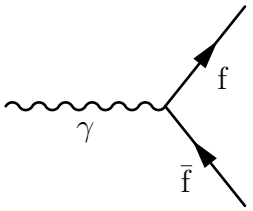


Figure 1.1: The basic QED vertex involving a γ and a charged fermion f.

It is convenient to use the Lagrangian formalism to describe the interactions of fermions within the SM, whereby such interactions can be described in terms of an action involving a Lagrangian acting on the fermion fields. The dynamics of non-interacting massive fermion fields is described by the Dirac equation [4] and as such the relevant Lagrangian is the Dirac Lagrangian. In 1954 Yang and Mills [5] proposed a framework for theories involving the exchange of vector bosons, such as those in QED. The Lagrangian in a Yang-Mills theory is invariant under transformations that are a function of space and time. This local gauge symmetry provides an accurate description of physical interactions and is therefore a desirable property of the SM theories. In order to ensure this symmetry is retained, a new massless vector field must be introduced which is identified as the photon field. This gives rise to the QED Lagrangian which is then observed to contain interaction terms, represented by processes such as in Figure 1.1. The requirement that the physical system remains invariant under local gauge transformations results in the conservation of electron charge, as confirmed by experiment. A local gauge transformation in QED can be represented as a transformation under the symmetry group U(1) corresponding to all unitary matrices of dimension 1×1. In this way the local gauge group of QED is called U(1). Using different symmetry groups, the same principle may be extended to the strong and weak interactions.

1.2.2 Quantum Chromodynamics (QCD)

Quantum Chromodynamics (QCD) is the quantum field theory describing the strong interactions. The strong force couples to colour charge, so only the coloured gluons and quarks are involved in strong interactions. The most basic QCD interaction vertex, involving the interaction of quarks (q) with a gluon (g) is shown in leftmost figure in Figure 1.2. The local gauge group of QCD is $SU(3)_C$ corresponding to the unitary group of 3×3 matrices with determinant 1. The three dimensional nature of this group is a consequence of there being three quark colours (C) and as such the quark fields transforming in the vector space of colour. To preserve local gauge invariance of the Dirac Lagrangian eight massless fields must be introduced which correspond to the eight gluons. These fields are also vector fields because the gluons have an intrinsic colour charge.

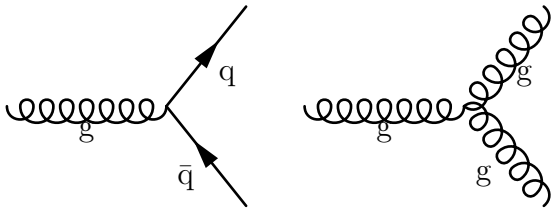


Figure 1.2: Basic QCD interaction vertices involving (left) gluon (g) and quark (q) anti-quark ($\bar{q}$) and (right) gluon self interaction.

The $SU(3)_C$ group is an example of a $non-abelian$ group, because its generators do not commute. This is a consequence of gluons carrying a colour charge and leads to self interaction terms in the QCD Lagrangian. This behaviour gives rise to diagrams such as those in the right of Figure 1.2 whereby gluons are coupling to other gluons. This does not occur in QED as photons do not carry electric charge. Two further consequences arise from this property, making QED and QCD intrinsically different. *Asymptotic freedom* refers to the fact that the strength of the strong force (i.e. the coupling strength) increases with increasing distance. This is because in QCD both quark anti-quark loops and gluon-gluon loops contribute to the higher order processes. However, although quark anti-quark loops (like loops in QED) lead to a net reduction in coupling strength with increasing distance the opposite is true for gluon-gluon loops. Because there are more gluons than quarks, the effect from the gluon-gluon loops outweighs that from the quark anti-quark loops and as such gives rise to a net increase in coupling strength with increasing distance. A consequence of asymptotic freedom is that no free quarks or gluons, unlike leptons and photons, are observed in nature and this is why we only observe colourless hadrons in the form of baryons or mesons.

1.2.3 The weak force and Electroweak theory

In analogy to QED and QCD, the quantum field theory of weak interactions is determined by requiring local gauge invariance of the appropriate Lagrangian. The local gauge group of the weak interaction under which the Lagrangian must be invariant is $SU(2)_L$ in the vector space of weak isospin I, where the L subscript refers to the fact that the fermions whose interactions the Lagrangian describes are left-handed (i.e. with intrinsic spin orientated opposite to direction of motion). All left-handed fermions experience the weak interaction and are arranged into pairs or fermion doublets. For the leptons these doublets consist of the same generations shown in Section 1.1, corresponding to the physical (mass) eigenstates. The weak interaction does not couple to the quark mass eigenstates (i.e. u,s,d) but instead, linear combinations of them (u′,s′,d′) which are determined by the CKM matrix [6]. This incorporates so called flavour changing charged current reactions, whereby

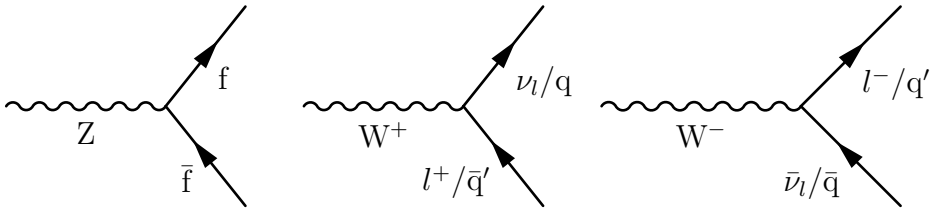


Figure 1.3: Weak interaction vertices allowed in the SM.

quarks of different flavours can be linked to a single weak vertex. The basic allowed weak interaction vertices are shown in Figure 1.3. Flavour changing neutral current reactions are not allowed in the SM. In the original formulation of the SM, neutrinos only experience the weak interaction and so are not predicted to have a right-handed component.

In 1968 Glashow, Salam and Weinberg [7, 8, 9] successfully extended the theory of weak interactions to encompass the electromagnetic interaction by using the gauge group $\mathrm{SU}(2)_L \bigotimes \mathrm{U}(1)_Y$. Here Y is called *weak hyper-charge* and is related to electric charge Q by $Q = Y + I_3$ where I_3 is the 3^{rd} component of weak isospin. This showed that the electromagnetic and weak forces can be viewed as two components of a single force, called the electroweak force at high energy. In this case preserving local gauge invariance requires four massless fields be introduced. Mixing of these four fields gives the electroweak bosons γ, W^+, W^- and Z. The non- abelian nature of the $\mathrm{SU}(2)_L$ group gives rise to self interaction terms and allows W(and Z) bosons to couple to each other. The $\mathrm{U}(1)_Y$ group is however *abelian* and as such the absence of photon-photon couplings in QED is maintained.

1.2.4 The Higgs mechanism

In the same way that the electromagnetic and weak theories were combined to form a unified electroweak theory, the theories of the electroweak and strong interactions may be combined to form a unified theory of all the fundamental forces apart from gravity. The corresponding local gauge group is $\mathrm{SU}(3)_C \bigotimes \mathrm{SU}(2)_L \bigotimes \mathrm{U}(1)_Y$ and the associated Lagrangian, the SM Lagrangian. However as with the individual theories, the fields which must be included in order to preserve local gauge symmetry (these are identified with the γ, W^+, W^-, Z and gluons) are each required to be massless. The same is true for the fermions which under $\mathrm{SU}(2)_L$ must be massless. Experimentally, however, it has been shown that the W^+, W^-, Z and fermions are indeed massive.

The mechanism of *spontaneous electroweak symmetry breaking* applied to a $non-abelian$ theory was introduced by Peter Higgs [10] and others [11, 12] in 1964 and provides a solution to the massless fields. This is what is commonly known as the Higgs mechanism.

The basic idea behind *spontaneous symmetry breaking* is discussed in the following by means of adding to the theory a complex scalar field $\phi = \frac{\phi_1 + i\phi_2}{\sqrt{2}}$ with Lagrangian Eqn. 1.1

$$\mathcal{L} = T - V(\phi) = (\partial_\nu \phi)(\partial^\nu \phi) - \frac{1}{2}\mu\phi^*\phi - \frac{1}{4}\lambda(\phi^*\phi)^2 \tag{1.1}$$

where $V(\phi)$ is the potential and μ and λ are two free parameters. In order for this Lagrangian to be invariant under global gauge transformations (the associated gauge group is U(1)), i.e. symmetric under $\phi \to -\phi$, and there to exist a vacuum state with positive and finite energy, λ must be positive. Imposing $\lambda > 0$, two solutions for μ^2 exist. When $\mu^2 > 0$, the potential has one minimum and as such the vacuum state is unique. If however, if $\mu^2 < 0$, the potential minima is now given by Eqn. 1.2

$$\phi_1^2 + \phi_2^2 = v^2 \tag{1.2}$$

where v corresponds to the vacuum expectation value of the field. The potential minima are now not unique and lie on a circle in the (ϕ_1,ϕ_2) plane. The potential V(ϕ) is shown in Figure 1.4 for both $\mu^2 < 0$ and $\mu^2 > 0$. By choosing the vacuum state corresponding to $< \phi_1 > = v$ and $< \phi_2 > = 0$, the global gauge symmetry is spontaneously broken. This is called spontaneous symmetry breaking.

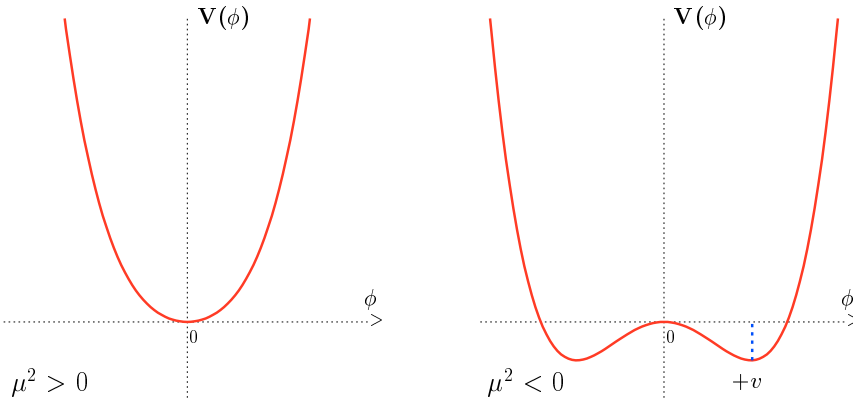


Figure 1.4: The Higgs Potential ($V(\phi)$) for the case with $\mu^2 > 0$ (left) and $\mu^2 < 0$ (right) [13].

To see how this gives rise to mass terms in the Lagrangian, one can expand ϕ around the vacuum state, giving Eqn. 1.3, thereby introducing two new fields η and ρ representing the quantum fluctuations around the vacuum state.

$$\phi(x) = \frac{1}{\sqrt{2}}(v + \eta(x) + i\rho(x)) \tag{1.3}$$

Rewriting the Lagrangian in Eqn. 1.1 using this redefined potential yields Eqn. 1.4.

$$\mathcal{L} = \frac{1}{2}(\partial\eta)^2 + \frac{1}{2}(\partial\rho)^2 - \frac{1}{2}\mu^2\eta^2 + \text{higher order terms} \tag{1.4}$$

This form of the Lagrangian contains a mass term (m) of the form $-\frac{1}{2}m^2\eta^2$ where $m = \sqrt{-2\mu^2}$. However, this choice of μ^2 also gives rise to a field ρ with no associated mass term, indicating the presence of a new massless boson. This corresponds to the Goldstone boson, which is absorbed within the longitudinal component of the Higgs field.

The application of the above requiring local gauge invariance under the U(1) gauge group of the Lagrangian of the complex scaler field refers to the Higgs mechanism. The Higgs mechanism was included into the SU(2) electroweak theory by Weinberg and Salam, who used a weak isospin doublet of complex scalar fields with hyper charge Y =1, called the Higgs doublet (Eqn. 1.5).

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{1.5}$$

This field transforms in the same way as electroweak doublet fields. Its Lagrangian includes covariant derivatives to ensure that local gauge invariance under $SU(2)_L \otimes U(1)_Y$ is preserved. Breaking the symmetry with $\mu^2 < 0$, gives rise to the degenerate vacuum state. Weinberg chose the vacuum state

$$< \phi >= \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix} \tag{1.6}$$

thus keeping the ground state electrically neutral and the photon massless. $v = \sqrt{-\frac{\mu^2}{\lambda}}$ is the vacuum expectation value of the potential. Expanding around the ground state as was done previously, the field is redefined as

$$\phi = e^{i\kappa_a(x)\beta_a/2v} \begin{pmatrix} 0 \\ \frac{v+\mathrm{H}(x)}{\sqrt{2}} \end{pmatrix} \Longrightarrow \phi = \begin{pmatrix} 0 \\ \frac{v+\mathrm{H}(x)}{\sqrt{2}} \end{pmatrix} \tag{1.7}$$

where κ are three real fields corresponding to Goldstone bosons, H is the Higgs field and β are the Pauli matrices. Since the SM Lagrangian is invariant under SU(2) transformations, by applying the SU(2) transformation $e^{i\kappa_a(x)\beta_a/2v}$ the form of the field can be simplified to the final expression in Eqn. 1.7.

Substituting this expression into the Lagrangian, yields mass terms from applying the covariant derivative to the field. From these the physical mass of the W,Z and γ fields are obtained as $\mathrm{m_W}$, $\mathrm{m_Z}$ and $\mathrm{m_A}$ respectively

$$\mathrm{m_W} = \frac{v g_w}{2}, \quad \mathrm{m_Z} = \frac{v\sqrt{g_w^2+g'^2}}{2}, \quad \mathrm{m_A} = 0 \tag{1.8}$$

where g_w and g' are coupling constants. The Higgs boson mass, as in Eqn. 1.9, is found using the potential.

$$\mathrm{m_H} = \sqrt{-2\mu^2} = v\sqrt{2\lambda} \tag{1.9}$$

Since λ is a free parameter the Higgs mass cannot be calculated but must be measured by experiment. Spontaneous symmetry breaking in this way does not generate masses for the fermions. Instead additional Lagrangian terms representing the interaction of the fermion fields with the Higgs field must be added, whilst preserving $SU(2)_L \otimes U(1)_Y$ gauge invariance. The form of the corresponding Lagrangian is shown in Eqn. 1.10

$$\mathcal{L} = -C_f[\bar{\Psi}_L^f \phi \Psi_R^f + \bar{\Psi}_R^f \phi^\dagger \Psi_L^f] \tag{1.10}$$

where $-C_f$ corresponds to the Yukawa coupling and Ψ^f the fermion field for the fermion f. As previously, the field can be expanded around the ground state and the symmetry spontaneously broken to yield mass and interaction terms by re-substituting into the Lagrangian. These have the form of

$$-\frac{C_f}{\sqrt{2}} v \bar{\Psi}_f \Psi_f \longrightarrow \text{mass term} \tag{1.11}$$

$$-\frac{C_f}{\sqrt{2}} v \bar{\Psi}_f \Psi_f H \longrightarrow \text{interaction term} \tag{1.12}$$

From this it is seen that all quarks and leptons couple to the Higgs in the same way. The strength of their coupling is proportional to their mass m_f and given by

$$C_f = \sqrt{2}\frac{m_f}{v} \tag{1.13}$$

1.3 Constraints on the Higgs boson mass

The Higgs boson mass is a free parameter in the SM and as such cannot be predicted. Despite this there exist a number of ways to constrain its mass broadly speaking coming from theoretical and experimental means. The experimental constraints on the Higgs bosons mass come from searching for its production, by colliding particles at high energy in particle colliders.

1.3.1 Theoretical constraints

The theoretical constraints on the Higgs boson mass are

Unitarity Unitarity is the requirement that the total scattering probability for a process, obtained from integrating over all contributing Feynman diagrams, remains less than unity. Fermi's theory of weak interactions violated unitarity at the electroweak scale $\sqrt{s} \approx G_\mu^{-1/2}$ because it assumed point interactions. Introduction of massive intermediate bosons resolved this problem for low energy. However, certain processes involving the longitudinal components of the vector bosons are expected to violate unitarity at tree-level. An example of such a process is $\mathrm{W^+W^-} \rightarrow \mathrm{W^+W^-}$ longitudinal scattering which gets contributions from Z and γ, leading to its cross section increasing at high energy in proportion to the square of the centre of mass energy ($\sqrt{s}$). Unitarity can be restored by adding Higgs exchange diagrams, but this places an upper bound on the Higgs mass of $2v\sqrt{\pi} \approx 800$ GeV/c^2.

Triviality As shown in Section 1.2.4, the mass of the Higgs boson is given by Eqn. 1.9. This represents the leading order expression and will be modified once higher order corrections are accounted for, examples of which are shown in Figure 1.5. These corrections give rise to divergences which can be accounted for using a renormalisation procedure. After applying this renormalisation it becomes evident that the Higgs self-coupling λ diverges with increasing energy scale. Assuming the Higgs self-coupling is larger than the top quark Yukawa coupling, it varies with energy like $\lambda^2 \ln(\mathrm{Q}^2)$. As such if it is assumed that the SM is valid at all energies, λ must be zero. This implies that the SM is valid up to a cut-off energy scale (Λ) at which new physics will begin to appear. In order that it produces meaningful predictions at energies below this cut-off the perturbativity of the SM theory must be maintained and as such the Higgs self coupling must remain finite. Since the Higgs coupling is proportional to the square of the Higgs mass, an upper limit on its

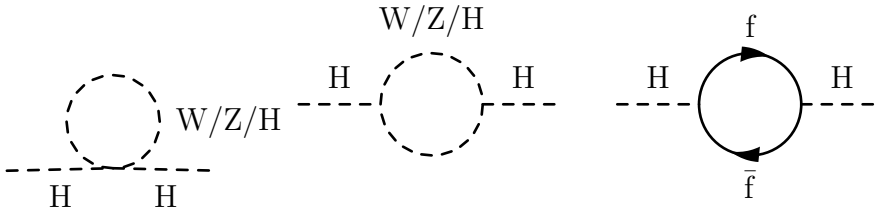


Figure 1.5: Feynman diagrams showing one-loop corrections to the SM Higgs boson mass, due to (left,middle) W,Z,H bosons and (right) fermions. W,Z and H bosons are represented by the dashed line and fermions by the solid line. In the left and middle plots the loops consist of W,Z,H bosons while the horizontal line represents the H boson.

value is predicted, depending on Λ. This is called the triviality bound on the mass of the Higgs boson. For a cut-off energy of 10 TeV the upper limit imposed by this constraint is $m_H \approx 500$ GeV/c^2.

Vacuum stability A lower constraint on the Higgs boson mass is derived by assuming the top quark Yukawa coupling is larger than the Higgs self-coupling. This implies low Higgs masses (i.e. $m_H < m_t$) where the coupling to top and weak bosons becomes large. In this scenario, the Higgs potential will develop a global minimum at large energy scales, thereby creating an unstable vacuum and preventing *spontaneous symmetry breaking*. Imposing the same cut-off energy scale (Λ) at which the SM is valid to, a lower constraint on the Higgs boson mass is imposed in order to maintain vacuum stability. For a cut-off energy scale of the electroweak scale ($\Lambda = 10^3$) $m_H \approx 70$ GeV/c^2. The combined effect of the triviality and vacuum stability requirements is shown in Figure 1.6, showing the allowed Higgs mass window as a function of the cut-off energy scale Λ.

1.3.2 Experimental constraints

In 1989 the Large Electron Position (LEP) collider became operational. It operated in two phases, called LEP-1 and LEP-2. The predominant Higgs production mechanism was through Higgs-strahlung i.e. $e^+e^- \rightarrow$ HZ. Data taken at LEP-1, when the collider was operated at a centre of mass energy close to the Z mass, excluded a Higgs mass $< 65.2\,\mathrm{GeV}/c^2$. During this running phase the most important decays were H$\rightarrow$ b$\bar{\text{b}}$ and Z $\rightarrow \ell\bar{\ell}$, where ℓ refers to a pair of charged leptons (except τ) or neutrinos. LEP-2, when the collider was operated at energies up to $209\,\mathrm{GeV}/c^2$, led to a lower limit on the Higgs boson mass of $m_H > 114\,\mathrm{GeV}/c^2$ being established to 95 % Confidence Level (C.L.). Two other decay modes, one involving a four jet final state with H $\rightarrow$ b$\bar{\text{b}}$ and Z $\rightarrow$ q$\bar{\text{q}}$ and another involving τ's H $\rightarrow \tau^+\tau^-$,b$\bar{\text{b}}$ Z $\rightarrow$ q$\bar{\text{q}}$,$\tau^+\tau^-$ became accessible and contributed to this result. The CDF and D0 Collaborations search for the SM Higgs boson using the Tevatron p$\bar{\text{p}}$ collider operating at $\sqrt{s} = 1.96$ TeV is ongoing. For $m_H < 130\,\mathrm{GeV}/c^2$, where Higgs decays to b$\bar{\text{b}}$ are considered, the most important mode of Higgs boson production is through

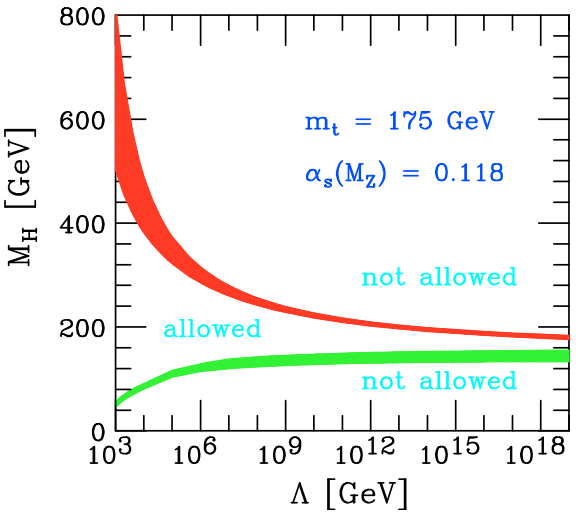


Figure 1.6: Theoretical upper and lower limits on the mass of the Higgs boson as a function of cut-off energy scale Λ. The upper limits are provided by the triviality bound and the lower limits by the vacuum stability bound. The bands represent the upper and lower limit theoretical uncertainties and enclose the allowed Higgs boson mass. The limits were derived assuming a top mass $m_t = 175 \pm 6\,\mathrm{GeV}/c^2$ and strong coupling constant $\alpha_s = 0.118 \pm 0.002$. [14].

associated production with a W or Z boson $q\bar{q} \rightarrow WH,ZH$, which allows for suppression against QCD background. Dominant for $m_H > 130\,\mathrm{GeV}/c^2$, gluon fusion $gg \rightarrow H$ and vector boson fusion $q\bar{q} \rightarrow q\bar{q}H$ are investigated with the Higgs boson decaying to $WW^* \rightarrow l\nu l\nu$. The amount of data used in a search is commonly represented in terms of an integrated luminosity corresponding to the sum over time of instantaneous luminosities. Instantaneous luminosity ($\mathcal{L}$) represents the particle flux giving rise to high energy collisions per unit area per unit time and as such integrated luminosities are expressed in terms of a unit area. The barn corresponding to 10^{-28} m^2 is most commonly used in particle physics. As of summer 2011, using the 8.6 fb^{-1} of data accumulated by both collaborations, the Tevatron experiments have excluded to 95 % C.L. a SM Higgs boson with mass $100 < m_H < 108\,\mathrm{GeV}/c^2$ (already excluded by the LEP experiments) and $156 < m_H < 177\,\mathrm{GeV}/c^2$ [15], as shown in Figure 1.7.

The LHC began operation in 2009 operating at $\sqrt{s} = 7\,\mathrm{TeV}$ and since then searches for the SM Higgs boson utilizing data have become established. As of summer 2011 the ATLAS Collaboration, using a combination of search channels each using between 1.0 to 2.3 fb^{-1} of data, produced preliminary results [16] indicating exclusion of the Higgs boson mass ranges from $146\,\mathrm{GeV}/c^2$ to $232\,\mathrm{GeV}/c^2$, 256 to $282\,\mathrm{GeV}/c^2$ and 296 to $466\,\mathrm{GeV}/c^2$ at the 95 % C.L., as shown in Figure 1.8(a). In the absence of a signal the expected Higgs boson mass exclusion ranges from 131 to $447\,\mathrm{GeV}/c^2$. The CMS Collaboration has also presented results in summer 2011 [17]

using a range of search channels. Using between 1.1-1.7 fb^{-1} of data they exclude the SM Higgs boson to 95 % C.L. in the mass ranges 145-216 GeV/c^2, 226-288 GeV/c^2 and 310-400 GeV/c^2 as shown in Figure 1.8(b).

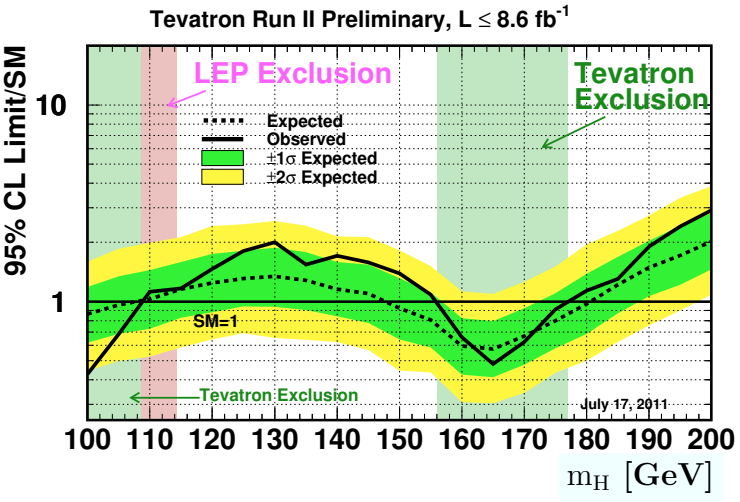


Figure 1.7: Summer 2011 Higgs combination limits from the CDF and D0 Collaborations with integrated luminosities up to 8.6 fb^{-1}. This excludes a SM Higgs boson at 95 % C.L. with mass $100 < \mathrm{m_H} < 108\,\mathrm{GeV}/c^2$ or $156 < \mathrm{m_H} < 177\,\mathrm{GeV}/c^2$ [15]. The combined upper limit on the SM Higgs boson production cross section divided by the SM expectation as observed is presented as a function of $\mathrm{m_H}$ by the solid line. The expected exclusion is shown by a dashed line.

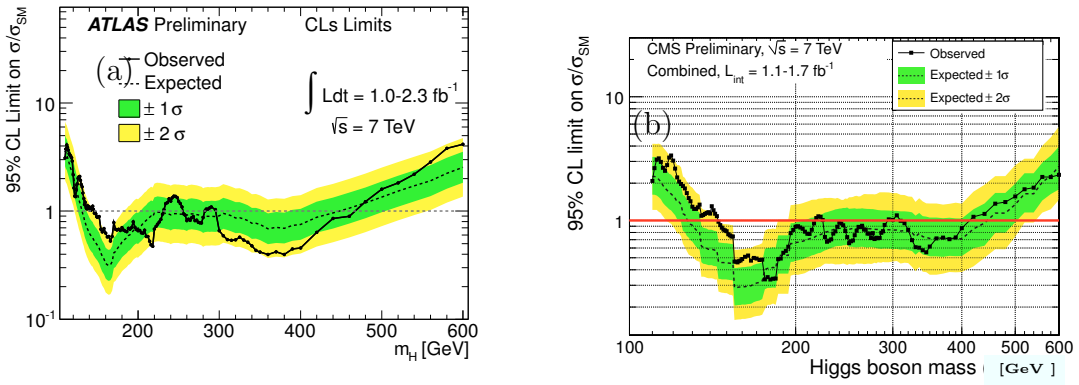


Figure 1.8: Exclusion limits from SM Higgs searches done by the (a) ATLAS [16] and (b) CMS [17] Collaborations as of summer 2011. In each case the combined upper limit on the SM Higgs boson production cross section divided by the SM expectation as observed is presented as a function of $\mathrm{m_H}$ by the solid line. The expected exclusion is shown by a dashed line.

Indirect experimental limits on the Higgs boson mass have been established using

precision measurements of electroweak parameters, predominantly measured at the Large Electron Positron (LEP), Stanford Linear and Tevatron colliders. Using the Higgs mass as a free parameter and fitting the SM prediction to the results from the precision measurements, shown in Figure 1.9, a Higgs mass of $m_H = 84^{+34}_{-26}\,\mathrm{GeV}/c^2$ is predicted. Inclusion of the results from the direct searches at LEP and the Tevatron (up until July 2010) in the fit, suggest $m_H = 120.6^{+17}_{-5.2}\,\mathrm{GeV}/c^2$.

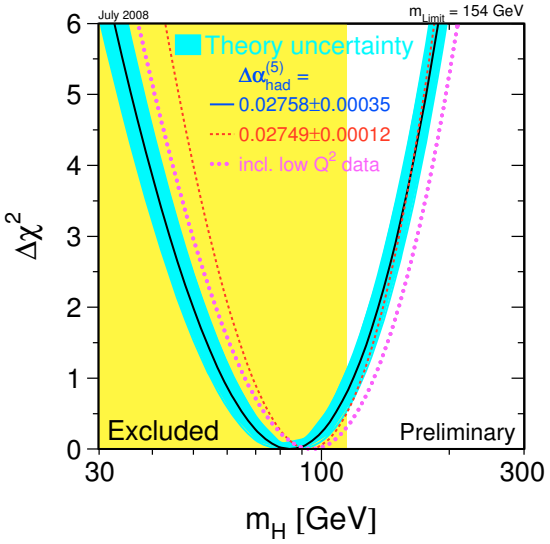


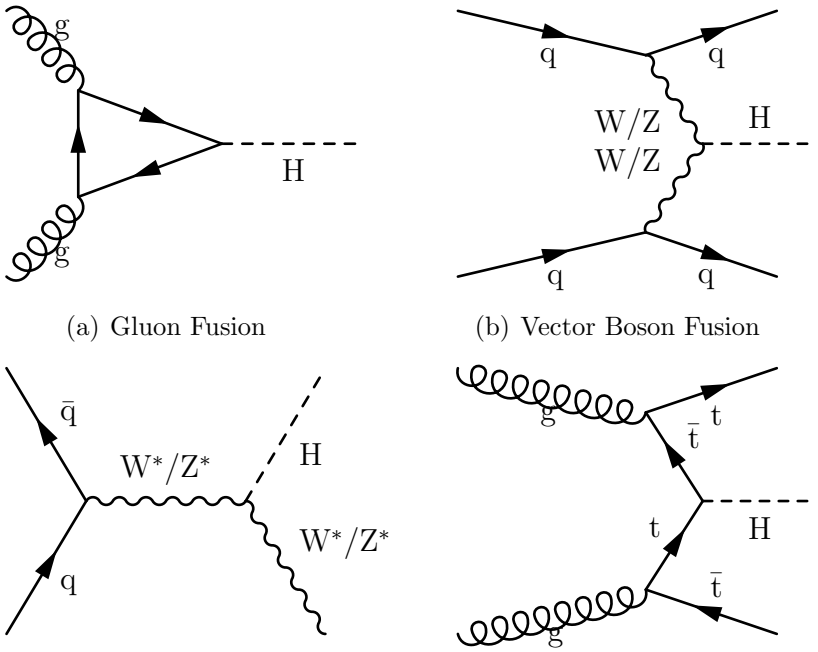
Figure 1.9: $\Delta\chi^2$ of the combined fit to the electroweak precision measurements as a function of Higgs boson mass (m_H) [18]. The theoretical uncertainty represented by the blue band corresponds to unknown higher order corrections. The regions shaded yellow represent the Higgs boson masses excluded at 95 % C.L. from direct Higgs searches at the Tevatron and LEP.

1.4 Higgs at the LHC

1.4.1 Higgs production mechanisms

Higgs production at the LHC can be divided into four main mechanisms, gluon gluon fusion (GF) ($gg \rightarrow H$), Vector Boson Fusion (VBF) ($q\bar{q} \rightarrow q\bar{q}H$), associated production with W/Z boson ($q\bar{q} \rightarrow HW/Z$) and associated production with heavy quarks ($gg/q\bar{q} \rightarrow q\bar{q}H$). Feynman diagrams showing each of these processes are shown in Figure 1.10. The corresponding cross section for each mechanism is shown in Figure 1.11 as a function of Higgs mass for $\sqrt{s} = 7\,\mathrm{TeV}$.

GF mediated by heavy quark loops is the dominate production mode. This is largely due to higher order QCD corrections, with next-to-leading order (NLO) effects increasing its total cross section by $\sim$ 80-100 % at the LHC [19].



(a) Gluon Fusion

(b) Vector Boson Fusion

(c) Associated Production W/Z

(d) Associated Heavy Quark Production

Figure 1.10: Diagrams of SM Higgs Production Modes at the LHC.

VBF is suppressed by approximately one order of magnitude compared to GF according to the SM. However this mode of production produces a Higgs boson in association with two quarks, leading to the production of two highly energetic jets typically in the forward regions of the detector with no jet activity other than that produced from the Higgs decay products between them (because of no colour flow between the initial interacting particles). This results in a distinctive signal signature allowing for efficient suppression of backgrounds.

The associated production modes have lower cross section than either GF or VBF. Nonetheless they will need to be studied in order to verify the validity of the SM prediction of Higgs production modes. In addition associated production with heavy quarks is likely to become important for measuring the properties of the Higgs boson, once it is discovered.

1.4.2 Decay modes of the Standard Model Higgs boson

The decay of the Higgs boson according to the SM can be loosely grouped into decays to fermion and gauge boson pairs (and virtual loops) (see Figure 1.12).

Branching ratios for such processes are known to NLO. At tree level, the coupling strength of the Higgs boson to fermions is proportional to the mass of the particles concerned. The net result is that a Higgs boson of a given mass will decay to the heaviest fermions that are kinematically accessible and as a consequence the decays of the Higgs boson can be further classified by Higgs mass. The branching ratios

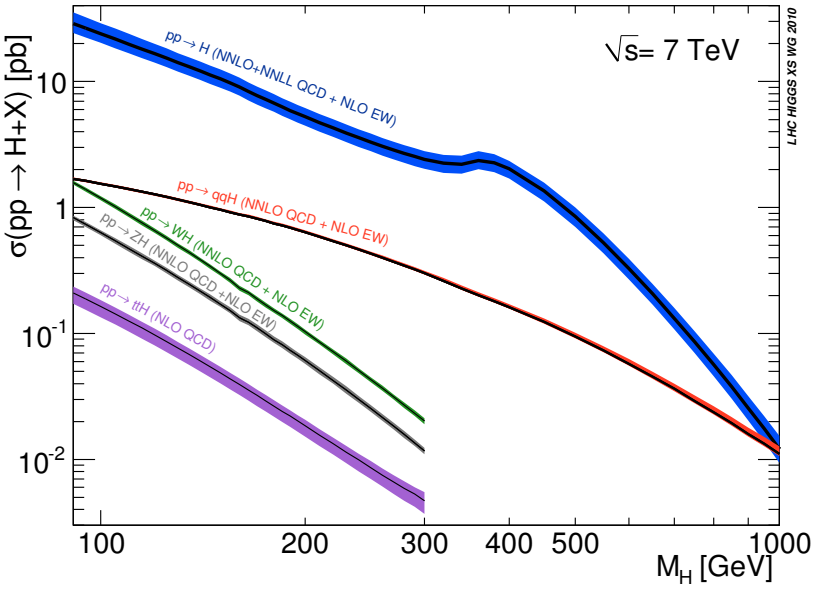
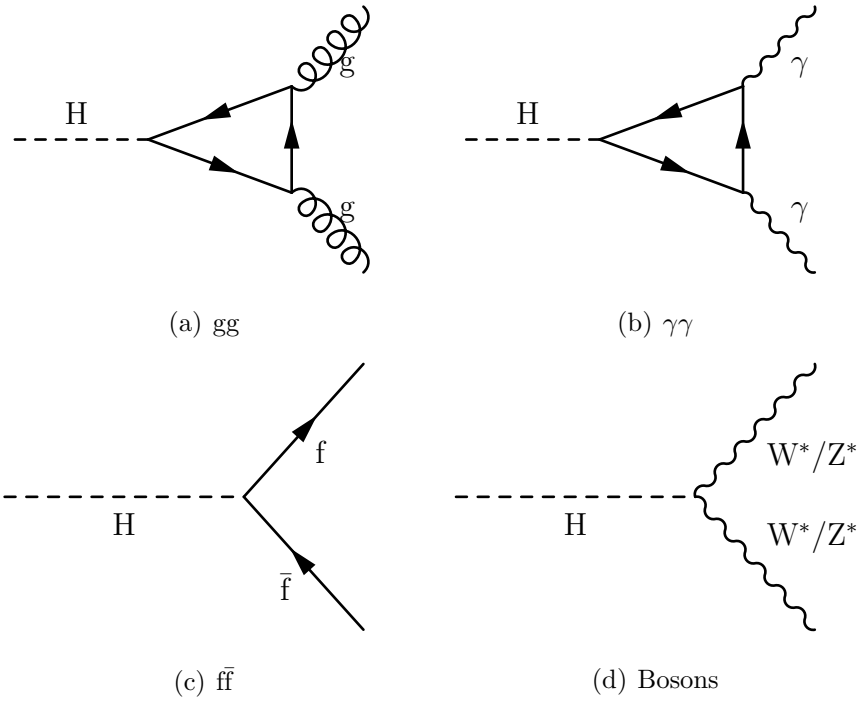


Figure 1.11: Production cross-sections for the SM Higgs boson as a function of Higgs boson mass for a LHC centre of mass energy $\sqrt{s} = 7$ TeV [20].



(a) gg (b) $\gamma\gamma$

(c) $f\bar{f}$ (d) Bosons

Figure 1.12: Diagrams of SM Higgs Decay Modes at the LHC.

(BR) of the SM Higgs boson for each decay mode, as a function of Higgs mass, are shown in Figure 1.13.

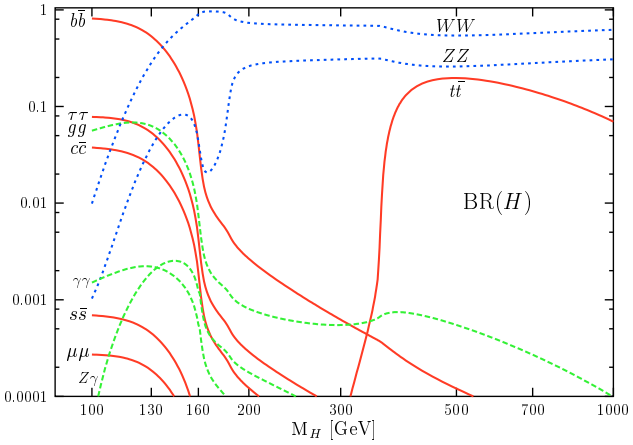


Figure 1.13: Branching ratios for the SM Higgs boson as a function of its mass [13].

$m_H < \mathbf{130}\,\mathrm{GeV}/c^2$ In this mass region H→ $b\bar{b}$ (BR ∼ 50-75 %) is the dominant decay mode. This channel suffers from extremely large background from jets in QCD processes and in order to achieve sensitivity requires very good sensitivity in measuring decays of b-hadrons. Observation through associated production with top quarks is also theoretically possible. H→ $\tau^+\tau^-$ (BR approx 5-7 %) has the next largest branching fraction. So that all τ decays can be considered, this mode is investigated in the vector boson fusion method of production. The other important decays in this mass range involve decays to massless γ's and g's, which therefore, do not couple directly to the Higgs boson. These decays occur via loops involving massive charged and/or coloured particles which couple to the Higgs boson. H→ $\gamma\gamma$ and H→ γZ are mediated by W boson and charged fermion (with only top quark contributing significantly for $m_H > 100\,\mathrm{GeV}/c^2$) loops. H→ gg is mediated by quark loops. Despite its comparatively low branching fraction, the striking signature it produces of two high p_T photons with invariant mass of the Higgs mass, means that H→ $\gamma\gamma$ is perhaps the best discovery channel for a low mass Higgs between the LEP limit and $m_H = 140\,\mathrm{GeV}/c^2$.

$\mathbf{130} < m_H < \mathbf{180}\,\mathrm{GeV}/c^2$ Decays to gauge boson pairs begin to dominate in this mass region. In the mass range $2m_W < m_H < 2M_Z$, H→ WW* is the main decay mode. As H → $\gamma\gamma$ is for the low mass region, H → ZZ* → $4l$ is the most promising discovery channel for a very large mass range, up to ∼ $600\,\mathrm{GeV}/c^2$.

$m_H > \mathbf{180}\,\mathrm{GeV}/c^2$ Here decays to gauge bosons are the dominant decay mode. $H \to WW$ has a BR $\approx 2/3$ while $H \to ZZ$ has a BR $\approx 1/3$ (since decays to ZZ are decays to same particle). Decays to $t\bar{t}$ start at $350\,\mathrm{GeV}/c^2$ contributing at most 20 %.

1.5 Beyond the Standard Model

Despite being able to explain with high precision most of the experimental data that has been produced to date, the SM suffers from several theoretical problems

- No dark matter candidate is provided by the SM.
- It doesn't explain the gravitational interaction.
- The level of CP violation is not sufficient to explain the matter anti-matter asymmetry seen in the universe.
- It doesn't explain the hierarchy problem, i.e. why gravity is so weak compared to the other interactions.
- Fine tuning is required to deal with divergences in the Higgs sector.

Because of these reasons and indeed others not discussed, it is a widely held opinion within the scientific community that the SM is an effective theory which we currently probe at low energy. The general theory will begin to become accessible when the predictions of the SM start to become incorrect.

More precise determination of the free parameters of the SM will allow the scale at which this happens to be better understood. Several beyond the SM theories exist which describe the SM predictions at low energy. Perhaps one of the most popular is supersymmetry which suggests a new symmetry between fermions and bosons and provides solutions to a number of the problems associated with SM. For example its models commonly provide a dark matter candidate and it also provides a solution to divergences in the Higgs sector.

Chapter 2

The LHC and the ATLAS Detector

This chapter gives an overview of the Large Hadron Collider [21, 22, 23, 24] and one of its main detectors ATLAS, the experiment on which this study is based. First a short introduction is made about the LHC focussing on its main parameters and design goals. Next, the ATLAS detector is described in terms of its major sub-detectors, their working technologies and design parameters.

2.1 The Large Hadron Collider (LHC)

The Large Hadron Collider (LHC) is a proton-proton (pp) collider with design centre of mass energy $\sqrt{s} = 14\,\text{TeV}$ and is the successor of the LEP collider. It has a 26.7 km circumference and is constructed approximately 100 m below ground level, on the Swiss-French border, installed in the existing tunnel used by the LEP collider. Its design characteristics are shown in Table 2.1.

Parameter	Value
Circumference	26.7 km
Beam injection energy	450 GeV
Design collision energy	14 TeV
Design instantaneous Luminosity	10^{34} cm^{-2}s^{-1}
Bunches per beam	2,808
Protons per bunch	1.15×10^{11}
Stored energy per beam	362 MJ
Bunch crossing rate	40 MHz
Time between bunch crossings	25 ns
RMS bunch length	7.55 cm
RMS bunch width	16.7 μm
Dipole temperature	1.9 K
Beam lifetime	22 h
Synchrotron Radiation loss per turn	7 keV

Table 2.1: Main LHC parameters.

Being a pp collider, the LHC's maximum $\sqrt{s}$ (like the Tevatron's) is not limited by synchrotron radiation (power emitted proportional to $1/m^4$ where m = beam particle mass) as was the case for the LEP collider. However, the production of antiprotons is highly inefficient so the LHC accelerates two counter rotating proton beams. This feature means that in contrast to $p\bar{p}$ colliders such as the Tevatron where both beams can share the same beam pipe, the LHC needs individual beam pipes for each beam, with opposite bending magnetic field orientations. Constraints as to the size of the accelerator, imposed by the diameter of LEP tunnel, meant that a twin-bore dipole magnet design, first proposed by J. Blewett [25], was adopted.

Altogether there are six experiments at the LHC, four of which ATLAS (A Toroidal LHC ApparatuS) [26], CMS (Compact Muon Solenoid), LHCb (LHC beauty) and ALICE (A Large Ion Collider Experiment) are located in dedicated caverns underground one at each of the four interaction points of the LHC. The two other experiments, LHCf and TOTEM are situated approximately 100 m from the interaction points of ATLAS and CMS respectively. ATLAS and CMS are the LHC's two multi-purpose experiments and are the central tools that will be used to fulfil the LHC's main physics objectives, including making precision measurements of the SM, finding evidence for a Higgs-like boson and exploring physics beyond the SM. However the other detectors also have important roles. LHCb will study B Physics and CP violation in the quark sector using b-hadrons. ALICE will search for evidence of quark-gluon plasma during LHC lead-ion collision runs. TOTEM will perform measurements of the pp cross section at the LHC while LHCf will study physics at small angles to the beam direction.

Before beams can be injected into the LHC, protons are extracted from an ionised hydrogen source and accelerated to 750 keV with Radio Frequency (RF) cavities. The second stage involves accelerating the protons further using a linear accelerator (LINAC 2) to increase the energy to 50 MeV. From here they are fed to the Proton Synchrotron Booster (PSB) and their energy is increased to 1.4 GeV. The two final steps involve passes through the Proton Synchrotron (PS) and then the Super Proton Synchrotron (SPS) where beam energy is increased to 26 GeV and 450 GeV respectively. Subsequently the beams are injected into the LHC where the 1232 superconducting dipole magnets keep them on circular trajectories and superconducting RF cavities operating at 400 MHz allow their energy to be further ramped by 485 keV per turn. Both the 1232 dipoles and the 392 quadrupoles used to focus the beams are operated at a temperature below 2 K using superfluid helium. The result is colliding beams that have a bunch structure with the protons localised periodically in space into bunches. A schematic layout of the CERN accelerator complex is shown in Figure 2.1.

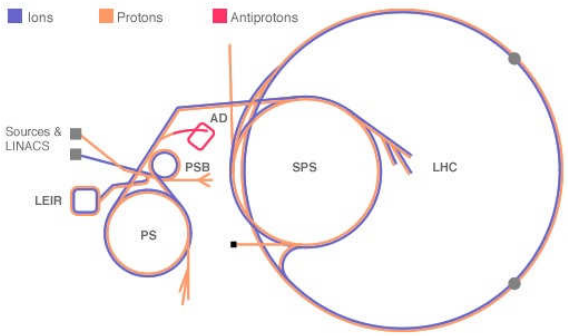


Figure 2.1: Overview of the CERN accelerator complex.

The rate of proton-proton interactions (called the event rate, R) caused by the collision of the counter-rotating proton beams at the LHC is proportional to the instantaneous luminosity $\mathcal{L}$ through $R = \mathcal{L}\sigma$, where σ corresponds to the event cross section or the probability of a particular interaction to occur (typically measured in barns as detailed in Chapter 1). The cross sections for a variety of processes at the Tevatron and the LHC are shown in Figure 2.2. The LHC has a design luminosity of $10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. Because increasing luminosity by increasing bunch crossing frequency puts high demands on experiment sub-detector electronic readout systems and trigger systems used to identify interesting processes, the high LHC luminosity will be achieved by increasing the density of protons per bunch, leading to multiple interactions per bunch crossing (called pile-up). When operated at its design luminosity, a nominal LHC beam will consist of approximately 2800 bunches spaced 25 ns apart and each containing around 10^{11} protons, giving an event rate of 40 MHz at each interaction point. This will lead to on average approximately 23 inelastic pp collisions per bunch crossing (pile-up events).

After successful first beam injection into the LHC during September 2008, technical problems occurred and resulted in the shut down of the LHC and postponement of colliding beams. Subsequently, first collisions at 900 GeV were achieved on November 23, 2009 and soon after the world record for the highest energy colliding beams was broken with 1.18 TeV beams giving $\sqrt{s} = 2.37$ TeV. By March 2010 the beams had been ramped to produce a first physics run at $\sqrt{s} = 7$ TeV, which is scheduled to continue during the majority of 2011 and provide 4-5 fb^{-1} of data for physics analyses. Collisions are planned to resume during 2012 at $\sqrt{s} = 7$ TeV prior to a pro-longed technical stop to allow for preparations for collisions at $\sqrt{s} = 14$ TeV starting 2014.

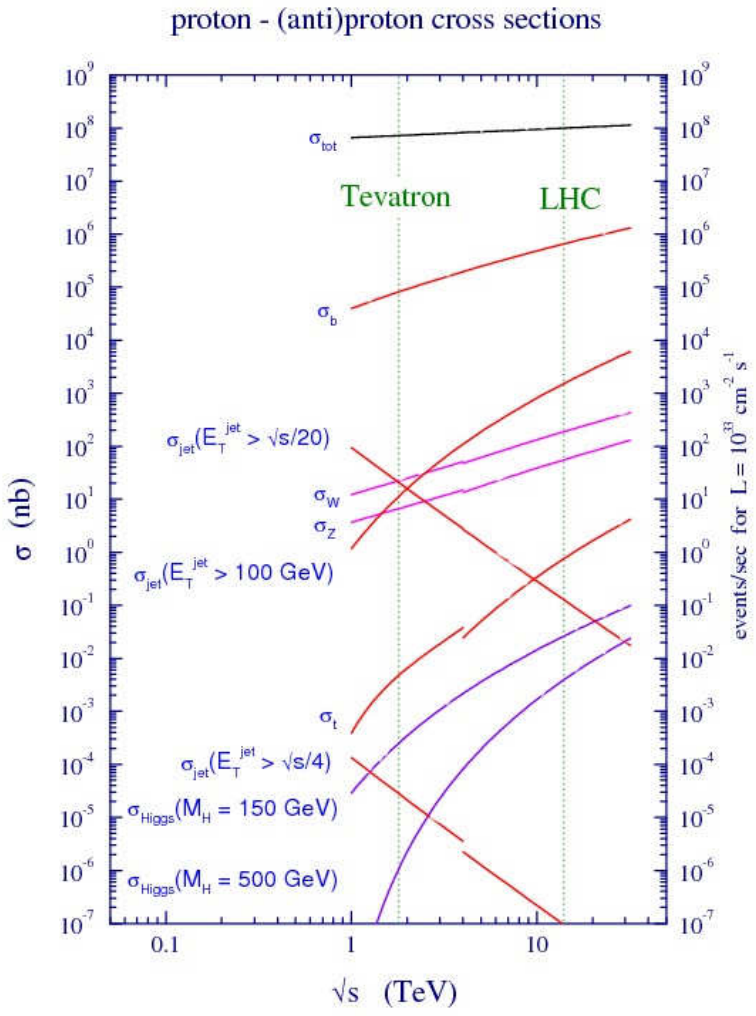


Figure 2.2: Production cross sections (nb) and event rates in proton-(anti)proton collisions as a function of centre-of-mass energy ($\sqrt{s}$) for an instantaneous luminosity of $10^{33}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ for different processes [27]. Discontinuities in curves at $\sqrt{s} = 4\,\mathrm{TeV}$ are due to the different colliding particles in the Tevatron (proton-anti-proton) compared to the LHC (proton-proton).

2.2 The ATLAS detector

ATLAS [28, 29] is one of the two general purpose detectors at the LHC. It is designed to detect the remnants of the high energy collisions produced by the LHC in order to test our current theoretical understanding of particle interactions. In order to do this, the remnants of the collisions must first be reconstructed into meaningful particle representations, from which the different physics processes of interest can be identified. Particle reconstruction in this context is explored in more detail in Chapter 4.

Arguably the primary goal of ATLAS (and CMS) is to establish the cause of spontaneous symmetry breaking in the electroweak sector, and as such its design has been guided and optimized to search for a Higgs-type boson. However the unprecedented energy and luminosity of the LHC means that ATLAS's physics programme also includes

- Precise SM measurements. Due to their large cross sections, W and Z bosons are produced copiously at the LHC (see Figure 2.2) and precision measurements of their properties have helped in the commissioning of ATLAS as well as providing a consistency check of the SM. The top quark was discovered at the Tevatron in 1995 and many of its properties studied. These will further be verified at the LHC, where the top-quark production cross section is more than two orders of magnitude higher than at the Tevatron. This is further motivated because top backgrounds will be dominant for many searches/measurements at the LHC.

- Beyond the SM: Several theories predict new physics at the TeV scale. As discussed in Chapter 1, Supersymmetry is a popular candidate. In addition, the W' and Z' bosons are examples of new particles predicted in this energy regime. Typically they decay to high p_T leptons.

A schematic layout of the ATLAS detector is shown in Figure 2.3, including each of its major sub-detectors. It shows ATLAS as having a layered composition with each detector sub-system arranged geometrically around the interaction point at the centre.

2.2.1 Nomenclature

The origin of the ATLAS co-ordinate system is defined to be the nominal interaction or collision point of the LHC's beams inside the ATLAS detector.

The positive z axis is defined by the trajectory of the clockwise (viewed from above) rotating proton beam. The xy (r) plane is transverse to this with the x axis

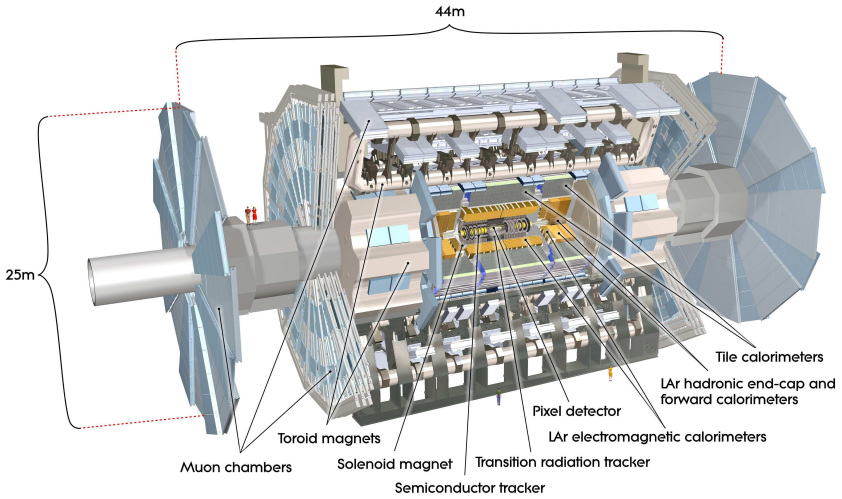


Figure 2.3: Schematic layout of the ATLAS detector showing the major sub-detectors [26].

pointing toward the centre of the LHC ring and the positive y axis pointing upwards. In this plane transverse variables such as transverse momentum $p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2} = p\sin(\theta)$ and transverse energy $E_{\mathrm{T}} = \sqrt{E^2 - p_z^2}$ are defined, where p_x, p_y, p_z are the x,y and z components of the particle's momentum and E is the particle's energy. θ is the polar angle measured from the z axis around the x axis and is often expressed in terms of pseudo-rapidity $\eta = -\ln(\tan\frac{\theta}{2})$, which equals the rapidity $y = \frac{1}{2}\ln\frac{E-p_z}{E+p_z}$ in the limit of small masses. Differences in rapidity are Lorentz-invariant under boosts along the z direction. The azimuthal angle $\phi = \tan^{-1}(p_y/p_x)$ is measured from the positive x axis clockwise around the z axis when facing the positive z direction. Typically distance in the $\eta - \phi$ plane is expressed in terms of $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

The physics goals of ATLAS impose the following design requirements:

- Radiation hard electronics to deal with the large particle fluence produced by the LHC.
- Full azimuthal coverage and large acceptance in η to maximize event reconstruction performance of low cross section processes.
- High detector granularity to allow efficient identification of particles created in the busy LHC environment.
- Efficient trigger to identify interesting events with low cross section and reject vast QCD background.
- Excellent identification and reconstruction of charged tracks including b-jets and τ-leptons through secondary vertices.

Detector component	Required Resolution	Coverage ($\|\eta\|$)
Tracking	$\sigma p_{\mathrm{T}}/p_{\mathrm{T}} = 0.05\,\% \; p_{\mathrm{T}} \oplus 1\,\%$	<2.5
Electromagnetic calorimetry	$\sigma_E/E = 10\,\%/\sqrt{E} \oplus 0.7\%$	<3.2
Hadronic calorimetry		
barrel/end-cap jets	$\sigma_E/E = 50\,\%/\sqrt{E} \oplus 3\,\%$	<3.2
forward jets	$\sigma_E/E = 100\,\%/\sqrt{E} \oplus 10\,\%$	3.1-4.9
Muon Chambers	$\sigma_{p_{\mathrm{T}}}/p_{\mathrm{T}} = 10\,\%$ at $p_{\mathrm{T}} = 1$ TeV	<2.7

Table 2.2: Performance goals of each component of the ATLAS detector including η coverage. The resolutions are expressed by two terms, the first representing statistical uncertainty and the second detector non-uniformity and calibration uncertainty. Units for p_{T} and E are in GeV unless stated otherwise.

- Good electromagnetic calorimetry for identification of electrons and photons. Hadronic calorimetry to perform precise jet measurements.
- Good muon identification and measurement of momentum up to 1 TeV.

In order to achieve these goals the different detector components were designed to meet certain energy/momentum resolution criteria. The required resolution of each of the detector components is summarised in Table 2.2.

2.2.2 Magnet system

To facilitate the measurement of particle momenta, ATLAS has a magnet system [30]. This consists of two main components i) a central solenoid surrounding the inner detector providing an axial B-field of 2 T across its volume and ii) a system of three air-core toroids encompassing the calorimeters providing a tangential B-field approximately orthogonal to muon trajectories. The magnet system is shown in Figure 2.4. The central solenoid is cylindrical in shape, 5.3 m in length and 2.4 m in diameter. To minimize the amount of material in front of the calorimeters it has a single layer coil structure made of superconducting NbTi coated in aluminium stabilizer and shares a cryostat with the barrel electromagnetic calorimeter. This means the total thickness, including support structure is 44 mm and results in a high magnetic stored energy to coil mass ratio of 7.4 kJ/kg at the nominal current of 20.5 kA.

The air toroid system is composed of three magnets: one forming a barrel section, and the others two end-caps. Each is composed of eight coils positioned evenly around the beam axis. The barrel is 25.3 m long and has an inner diameter of 9.4 m and outer diameter of 20.1 m. It provides a magnetic field strength of up to 3.9 T for the $|\eta| < 1.4$ region. The end-cap toroids are located either side of the central solenoid inside the barrel toroid. Each end-cap is rotated by 22.5 ° about the beam axis relative to the orientation of the barrel toroids to maximise the bending power

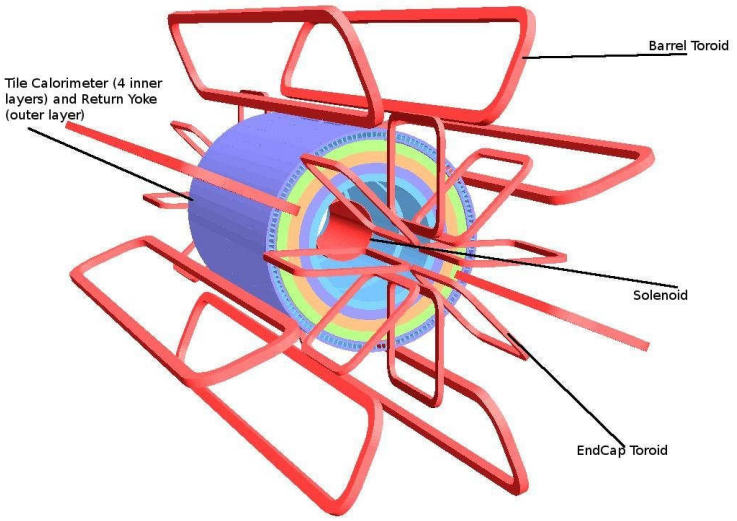


Figure 2.4: Geometry of magnet windings and tile calorimeter steel. The eight barrel toroid coils, with the end-cap coils interleaved are visible. The solenoid winding lies inside the calorimeter volume. The tile calorimeter is modelled by four layers with different magnetic properties, plus an outside return yoke [26].

in the transition region between the end-caps and barrel. The end-caps supply a magnetic field in the $1.6 < |\eta| < 2.7$ region with a maximum strength of 4.1 T. Conductor and coil winding technology is the same for the barrel and end-caps as that used in the central solenoid.

2.2.3 The inner detector

The inner detector's [31, 32] purpose is to accurately reconstruct charged tracks in and around the interaction point of ATLAS, which will see of the order to 1000 tracks/25 ns. More specifically, it must provide robust pattern recognition, measure track momentum and make primary and secondary vertex measurements (in order to enable identification of jets associated with decays of b hadrons, b-tagged jets). The inner detector is cylindrical in shape with overall length 7.2 m and diameter of 2.2 m providing $|\eta| < 2.5$ pseudo-rapidity coverage. The entire inner detector is composed of three separate but complementary sub-detectors, arranged concentrically around the beam pipe. The two innermost sub-detectors utilizing silicon technology are first the silicon pixel detector and then the SemiConductor Tracker (SCT). The outermost sub-detector is the transition radiation tracker (TRT). The overall structure of the inner detector is shown in Figure 2.5.

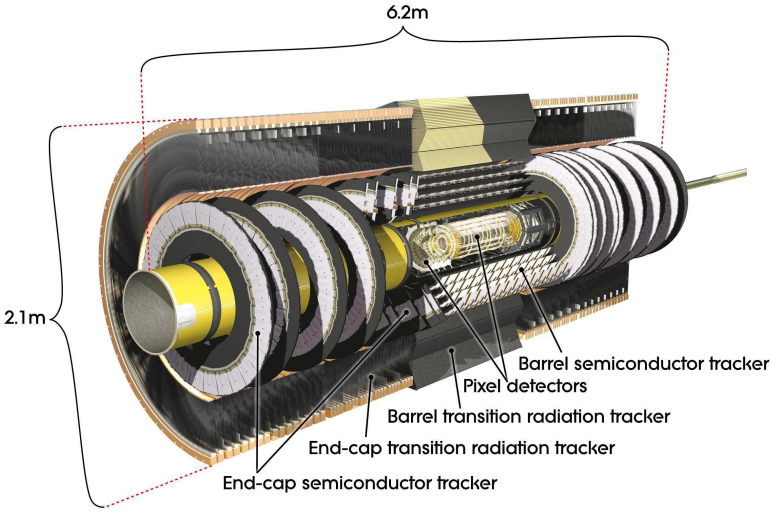


Figure 2.5: Picture of the ATLAS inner detector showing the 3 component sub-detectors, the Pixel, SCT and TRT [26].

Pixel Detector

The pixel detector is the innermost ATLAS sub-detector. It is composed of one barrel and two end-cap sections, each made up of three layers. In the barrel, the layers are positioned at radii of 5.05, 8.85 and 12.25 cm from the z axis. The end-cap layers are positioned perpendicularly to the beam axis at distances of 49.5, 58.0 and 65.0 cm from the nominal interaction point. Each layer is formed of modules, consisting of pixel detectors made of 1744 individual pixel sensors of dimension $50\times400\,\mu$m connected to 16 front end chips with 2880 readout channels. A charged particle traversing the pixel leads to the creation of electron-hole pairs and the applied voltage bias makes them move toward the readout cathode which measures the electrical signal produced. A hit is recorded if the signal exceeds a threshold value, whose position is determined by charge interpolation over adjacent pixels. The pixel detector has a measurement resolution of $10\,\mu$m in the $r-\phi$ plane and $115\,\mu$m in z.

Silicon Tracker

The silicon tracker is a semiconductor detector using the same detection principle as the pixel detector. It is composed of modules, each formed of two 768-strip $80\,\mu$m pitch sensors. The sensors are positioned at an angle of 40 mrad with respect to each other to provide measurement in the z direction. This feature also allows a measurement in R-ϕ when one of the two sensors is non-operable. Each module is read out with 1536 channels. The SCT modules are installed on a light support structure in four concentric barrel layers and two end-caps each consisting of nine

disks orientated perpendicular to the beam axis. The four barrel layers are comprised of 2112 rectangular SCT modules orientated 11 ° to the tangent of ϕ direction. This results in overlaps between modules that help to ensure the SCT can be properly aligned and that no dead regions in its coverage exist. The barrel covers the region $|\eta| < 1.4$. The position resolution of the SCT is 580 μm in the z direction and 17 μm in the $r - \phi$ plane.

Transition Radiation Tracker

The transition radiation tracker is composed of 4 mm diameter drift (straw) tubes filled with a Xe:CO_2:O_2 (70 %:27 %:3 %) gas mixture under 5-10 mbar pressure. 31 μm diameter tungsten wires plated with gold placed inside each of the straw tubes form the anode while the straws themselves form the cathode. A charged particle traversing a straw ionises the gas mixture inside and the resulting charge drifts toward the anode. Measurement of the drift time can be converted into a position. A position accuracy of 130 μm in the r-ϕ plane for $|\eta| < 2$ is achieved.

The TRT barrel is made up of three layers of 144 cm long straw tubes orientated along the z axis. The straw tubes are separated with polypropylene fibres. The end-caps are constructed from wheels. A wheel is made of eight layers of 768 radially orientated end-cap straws (37 cm long), with the space between the layers being filled with polypropylene fibres. Two different types of wheel are formed, one in which the layers are separated by 8 mm and another type with layer separation of 15 mm. Each end-cap is constructed of first 12 wheels with the smaller layer separation and then eight wheels constructed with the larger layer separation. Because polypropylene has a different dielectric constant to the straws, transition radiation in the form of photons is emitted as an incident charged particle traverses the boundary between them, contributing to the amount of ionisation it creates. The level of emitted transition radiation is dependent on the mass of incident particle through ($\gamma \propto E/m$), meaning a light particle will generate more transition radiation and consequently charge, than a heavier particle of the same energy. As such, two thresholds for the drift time measurement are implemented in the TRT front end electronics, allowing separation of charged particle tracks (low threshold) typically produced by high mass particles and charged particle tracks with transition radiation (high threshold) produced by lighter particles (electrons). Despite being less precise than the other inner detector sub-detectors, the TRT's large external radius of 1.08 m allows for many measurements per track, helping to improve the robustness of the inner detector momentum measurements.

2.2.4 Calorimetry

ATLAS calorimetry is designed to measure the energy of electrons, photons and hadrons. It is divided into three main parts, the electromagnetic calorimeter (ECAL), the hadronic calorimeter (HCAL) and the forward calorimeters (FCAL). They utilize different detector technologies to fulfil good energy and position resolutions and provide large hermetic coverage required for measurement of missing energy ($E_{\mathrm{T}}^{\mathrm{miss}}$), resulting when weakly interacting (ν) particles escape the detector undetected. In addition, they minimize hadron punch-through to the muon system. Each of the ATLAS calorimeters is a sampling calorimeter, i.e. they periodically sample or measure the energy of a traversing particle. To achieve this, each is composed of alternating layers of a dense absorber medium which causes traversing particles to shower and a sampling medium used to measure the energy of the resulting showers. The overall calorimeter system has a cylindrical construction and is positioned around the inner detector. It is approximately 13 m in length with a radius of 4.25 m, providing coverage for $|\eta| < 4.9$. The calorimeter system is shown in Figure 2.6.

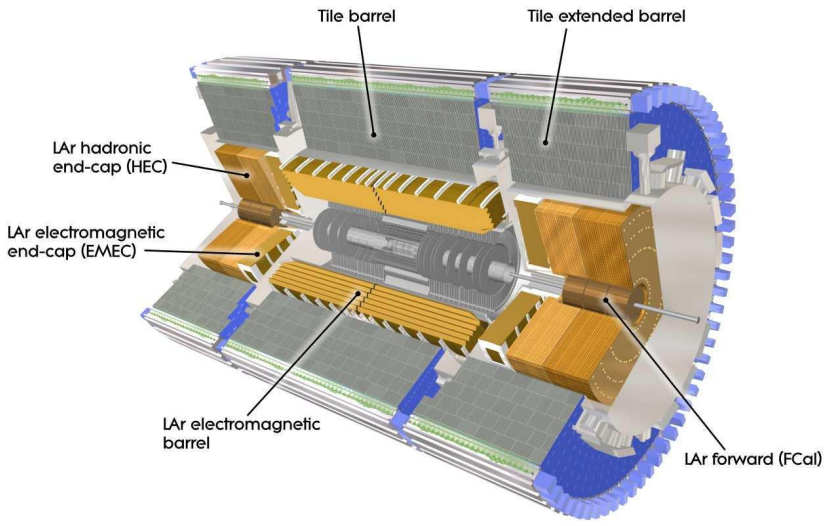


Figure 2.6: Picture of the ATLAS electromagnetic and hadronic calorimeters (cut-away view). [26].

Electromagnetic Calorimeter

The electromagnetic calorimeter [33] measures the energies of electrons and photons. Lead absorber material causes incident high energy electrons and photons to shower, by alternately pair production ($\gamma \rightarrow e^{+}e^{-}$) followed by bremsstrahlung ($e \rightarrow \gamma e$) [34]. Liquid argon (LAr) used as the sampling medium is then ionised

by the electrons/positrons and the signal produced from this is read out at copper electrodes by application of a voltage.

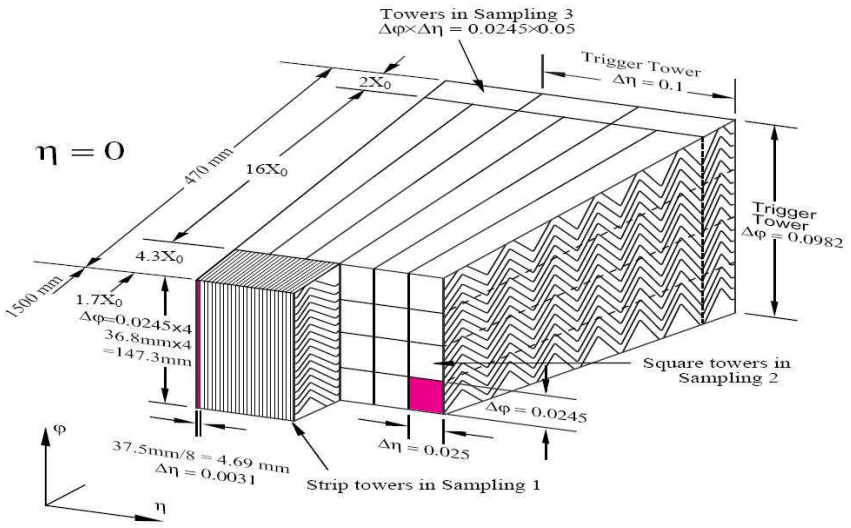


Figure 2.7: Diagram of a component of the ATLAS electromagnetic calorimeter, showing the three different samplings and the accordion geometry [26].

The electromagnetic calorimeter is made of two identical barrel sections and two end-caps, with a combined coverage up to $|\eta| < 3.2$. The barrel sections have an inner/outer radius of 1.4m/2.25 m while the end-caps are positioned perpendicular to the beam direction and are located between 33 cm and 2 m. Depending on η, each component of the ECAL is divided into two or more concentric layers called samplings, distinguished by their different granularity. The layout of these samplings is shown in Figure 2.7. In the region $|\eta| < 2.5$, covered by the inner detector, there are three samplings while for the region $2.2 < |\eta| < 3.2$ and the overlap region between the barrel ($|\eta| < 1.475$) and end-cap coverage ($|\eta| > 1.375$) there are two samplings. Each sampling is further divided into $\eta \times \phi$ cells whose dimensions are projective with respect to the nominal interaction point, varying with η.

The innermost sampling has the highest granularity, composed of strips of cells with dimension 0.003 $\times$ 0.1 in $\eta \times \phi$ at $\eta = 0$. This allows it to measure shower substructure and as such helps in identification of $\pi^0 \to \gamma\gamma$. It has a thickness of approximately $4\,X_0$, where X_0 is defined as the radiation length and characterises the length over which an high energy electron will lose $1/e$ of its energy by bremsstrahlung. The middle sampling has a much greater radial thickness of $17\,X_0$, and as such can be visualised as being made of towers of cells, each of size 0.025 $\times$ 0.025 in $\eta \times \phi$ (4×4 cm at $\eta = 0$). This cell size was chosen because it provides good containment of showers in a small number of cells which limits electronic and pile-up noise and because it allows discrimination of electromagnetic showers produced by

electrons and those from hadrons. The outer sampling has a cell size of 0.05×0.025 $\eta \times \phi$ with a thickness varying between 2 and $12\,X_0$.

In order to correct for energy losses of electrons and photons caused by detector material upstream of the calorimeter, pre-sampler layers of active LAr are positioned in front of the barrel ECAL covering $|\eta|$ <1.54 and between the barrel and end-cap cryo-stats covering $1.5 < |\eta| < 1.8$.

Hadronic Calorimeter

The purpose of the hadronic calorimeter [35] is to measure the energy and direction of jets from hadrons as well as assist in the measurement of $E_{\mathrm{T}}^{\mathrm{miss}}$ and prevention of punch-through to the muon system. Hadrons lose energy by interacting with the detector medium. This primarily occurs through inelastic interactions which result in the production of secondary strongly interacting particles, giving rise to hadronic showers. The distance a hadronic shower propagates is characterised by the interaction length (λ) and is commonly much larger than X_0 (i.e. hadronic showers propagate further and wider than electromagnetic showers). The hadronic calorimeter is divided into a barrel and two end-cap sections. The barrel is called the tile calorimeter (TileCal) and covers $|\eta| < 1.7$. The end-caps, together making the hadronic end-cap calorimeter or HEC cover $1.5 < |\eta| < 3.2$.

The hadronic end-cap calorimeter is a LAr sampling calorimeter using parallel copper plates as the absorber material. Each end-cap is made of two wheel structures each divided radially into two layers. The innermost layer of each wheel is constructed using 24 25 mm thick copper plates aligned perpendicular to the beam axis. The outer layers consist of 16 50 mm copper plates. The plates are separated by 8.5 mm gaps filled with LAr. For $1.5 < |\eta| < 2.5$ each layer has a cell granularity of 0.1×0.1 in $\eta \times \phi$ and for $2.5 < |\eta| < 3.2$, 0.1×0.1 in $\eta \times \phi$.

The tile calorimeter is composed of one 5.8 m long central and two 2.6 m long extended barrels covering $|\eta| < 1.0$ and $0.8 < |\eta| < 1.8$ respectively. Each is constructed from alternating layers of steel absorber and scintillating tile sampling material, with the gap between each barrel section for services filled with scintillating material. The tile calorimeter is longitudinally segmented into three layers with thicknesses of 1.5, 4.1 and $1.8\,\lambda$ in the central barrel and 1.5, 2.6, and $3.3\,\lambda$ in the extended barrels. Each layer is segmented in $\eta \times \phi$, with the two innermost layers having a granularity of 0.1×0.1 and the outermost 0.1×0.2.

Forward Calorimeter

The forward calorimeter's (FCal) purpose is to facilitate measurement of forward jets and improve the overall calorimeter hermeticity to add to the precision of $E_{\mathrm{T}}^{\mathrm{miss}}$ mea-

surement. It is made of two end-caps, located outside the electronic and hadronic calorimeter cryostats and covers $3.1 < |\eta| < 4.9$. Each endcap is built from three longitudinal modules forming three longitudinal layers. Each module is a LAr sampling calorimeter with the innermost module (for electromagnetic measurements), using copper and the outer modules (for hadronic measurements) tungsten for the absorber material. Each module is composed of a metal matrix. This method of construction ensures that gaps between the absorber layers can be kept very small (as low as 0.25 mm) in order to make the system tolerant to the high particle fluxes expected in this region.

2.2.5 Muon spectrometer

The purpose of the ATLAS muon spectrometer [36] is to provide trigger and bunch crossing identification of events with high p_{T} muons as well as high precision standalone momentum and position measurement of muons. Muons are more massive than electrons so the probability to interact via bremsstrahlung is lower. This coupled with the fact that they have a relatively long lifetime, means that the muons produced in LHC collisions will typically traverse the entire detector. For this reason the components of the muon spectrometer are located outside the calorimeters amongst the toroid magnet system in chambers. The overall system is comprised of a barrel divided into three concentric layers positioned at 5,7.5 and 10 m respectively and two end-caps each with four layers located at $7.4, 10.8, 14$ and 21.5 m either side of the interaction point parallel to the z axis. The barrel covers $|\eta| < 1.0$ and the end-caps cover $1.0 < |\eta| < 2.7$. Gaps in the muon spectrometer coverage occur at $|\eta| = 0$ to provide access and services to the inner detector, calorimeter and solenoid magnet and at ϕ= 1.2 and 2.2 because of ATLAS's support legs. The components/ chambers of the muon spectrometer are shown in Figure 2.8 including the toroid magnets that provide a means by which the trajectory of muons are bent in the $\eta - \phi$ plane.

In total four types of muon detection are used: chambers consisting of Monitored Drift Tubes (MDTs) and Cathode Strip Chambers (CSCs) which give precision measurements, and Thin Gap Chambers (TGCs) and Resistive Plate Chambers (RPCs) that provide triggering capability and a secondary position measurement in the direction orthogonal to that measured by the MDTs.

Precision Chambers

MDTs are used in both the barrel and end-caps, providing coverage for $|\eta| < 2.7$. They are used in all layers of the muon spectrometer except the innermost where because of their high rate ability CSCs are used. MDTs are constructed from 30 mm

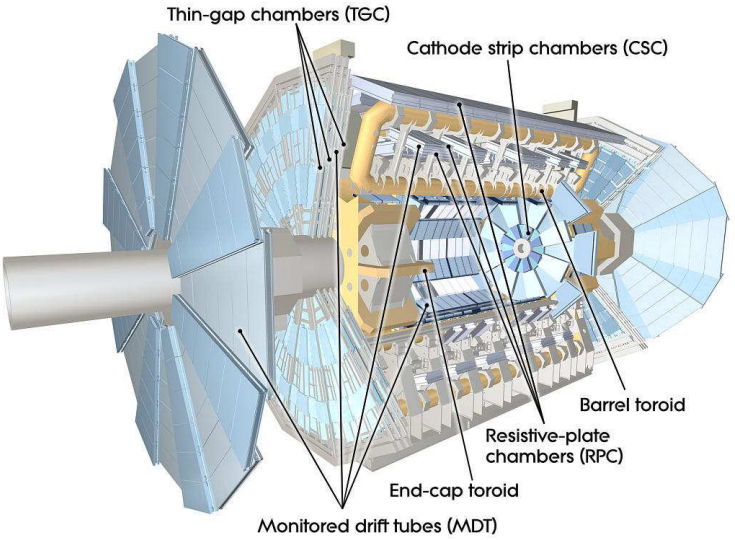


Figure 2.8: Picture of the ATLAS Muon detection system [26].

diameter 70-630 cm long aluminium tubes filled with Ar:CO_2 (93 %:7 %) with a central 50 μm tungsten-rhenium wire to generate an electric field in the gas. They are arranged into chambers. An incident charged particle ionises the gas and the electrons produced are drifted toward the anode wire where the charge is collected. In the innermost detector layer, each chamber is made of two layers of four layers of tubes while in the two outermost detector layers each chamber is made of two layers of three layers of tubes. The position resolution per tube is 80 μm and per chamber it is 35 μm.

CSCs are multi-wire proportional chambers used in the innermost layer of the end-caps, covering $2.0 < |\eta| < 2.7$. They consist of tubes filled with a mixture of Argon, CO_2 and CH_4 and central anode wires, perpendicular to which are the cathode strips. Their use in this high particle flux forward region is motivated because of their high counting rate capability, good spatial resolution and short electron drift time of less than 40 ns which allows them to operate in the high background. A charged particle will ionise the gas which is then measured at both the cathode and the anode, providing measurement of both r and ϕ coordinates. The resolution of the CSCs is 5 mm in ϕ and 40 mm in r.

Trigger Chambers

The ATLAS trigger system is designed to allow the efficient detection of interesting physics events. In addition it provides the capability of organising these events by associating them with a specific bunch crossing, from which they were derived. It is important to be able to trigger such events using muons as they are often contained in final states of interest. To this end, dedicated muon trigger chambers are used

which are designed to have a fast enough response time to be appropriate for this use case.

RPCs are composed of two parallel resistive bakelite plates separated by a 2 mm gap which is filled with a $C_2H_2F_4$, Iso-C_4H_{10}, SF_6 gas mixture. A potential difference is applied across the gap causing incident charged particles to give rise to ion showers within the gas. This signal is then read out on the outer sides of the resistive plates using metallic strips. RPCs provide barrel coverage for $|\eta| < 1.05$. RPCs are located on the same support structures as the MDTs and have the same dimensions. They are positioned so that a muon coming from the interaction point with sufficient energy will traverse three MDT and three RPC layers, allowing the measurement of position, in η and ϕ, in six places. RPCs have a spatial resolution of approximately 10 mm and can respond at time scales of the order of 1.5 ns.

TGCs cover the barrel section between $1.05 < |\eta| < 2.4$. Like CSCs they are multi-wire proportional chambers but the gap between the wire and cathode is smaller than the gap between the anode and the wire (1.4 mm as opposed to 1.8 mm). The gas mixture used is CO_2 with n-pentane (n-C_5H_{12}). TGCs have a spatial resolution of 2-6 mm in r and 3-7 mm in ϕ and a response time of around 4 ns.

2.2.6 Trigger and data aquisition

The design luminosity of the LHC ($\mathcal{L} = 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$) will give rise to 40 million bunch crossings occurring each second. It is expected that the total event rate after accounting for multiple interactions per bunch crossing will be of the order of 1 GHz, with a typical event size of the order of 1.5 MB. Technological limits currently place a restriction on the speed of recording events to disk at the level of ≈ 300 MB/s, thereby restricting the maximum rate of storing events to approximately 200 Hz. The ATLAS Trigger and Data aquisition (TDAQ) system is designed to facilitate the reduction of the event rate from the raw value of 1 GHz to 200 Hz and in doing so retain as many of the "interesting" physics events (i.e. those relating to the goals of ATLAS outlined in Section 2.2) as is possible. The Trigger system is based on three levels: Level 1, Level 2 and Event Filter, as illustrated in Figure 2.9. Level 2 and Event Filter together constitute what is termed the High Level Trigger (HLT) [37].

Level 1

Level 1 [39] is hardware-based and uses information from the calorimetry and the muon spectrometer. It is designed to reduce the 40 MHz bunch-crossing rate to 75 kHz. The system has a latency of 2.5 μs so pipeline memories are used to ensure a decision for each bunch crossing (occuring every 25 ns) is made within this time.

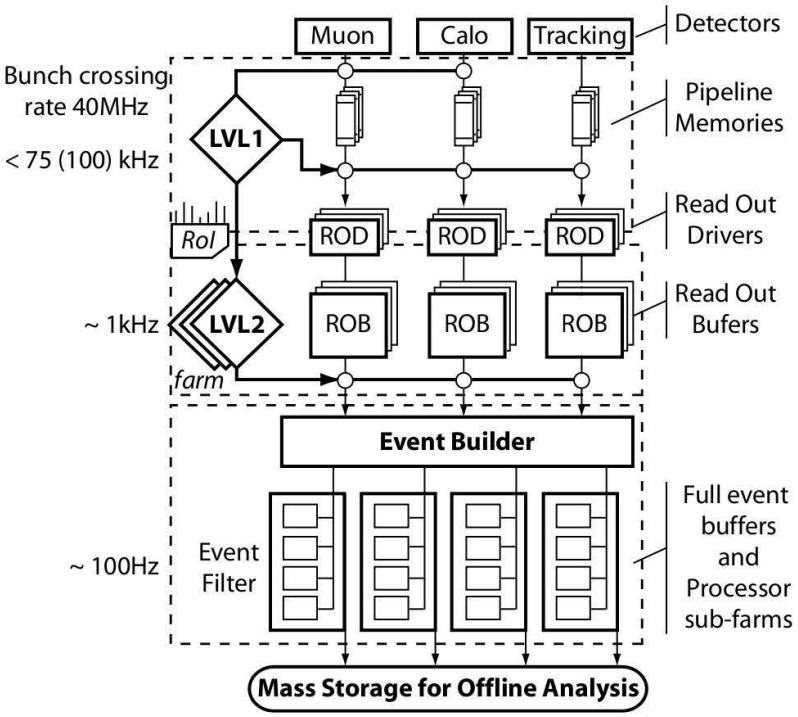


Figure 2.9: Overview of ATLAS Triggering system [38].

One of the tasks of Level 1 is to identify high p_{T} electrons/photons, muons, jets and τ leptons. It also selects events with large $E_{\mathrm{T}}^{\mathrm{miss}}$ and/or total energy.

Level 1 uses reduced granularity information from the RPCs and TGCs (as discussed in Section 2.2.5) for measuring high p_{T} muons. A set of predefined thresholds is applied which must be passed for a muon to be identified and triggered on. The low p_{T} thresholds require a p_{T} of between 6-9GeV/c whereas the high p_{T} thresholds require a p_{T} in the range 9-35 GeV/c. The thresholds must be passed in coincidence in 2(3) stations of chambers for the low(high) p_{T} thresholds, in order to reduce backgrounds from cosmic rays.

All the calorimeter sub-systems are used to identify electromagnetic clusters associated to electrons, photons, jets and τ leptons, as well as events with large $E_{\mathrm{T}}^{\mathrm{miss}}$ or large total transverse energy. This identification procedure is performed by level 1 algorithms which use information from calorimeter trigger towers. Calorimeter trigger towers are formed from summing the energy in all cells within 0.1×0.1 ($\eta\times\phi$) radial towers through all calorimeter sampling layers. To be identified, an object must pass a predefined p_{T} threshold. The readout of the data from the calorimeter and the muon spectrometer is performed by Read Out Drivers (ROD's).

The Central Trigger Processor (CTP) uses information from the calorimeter and muon triggers to establish, based on comparison with sets of predefined selections, whether each event should be considered further. If this criteria is met, the kinematic

information of each identified object is passed to the read out buffers in the form of Regions of Interest (ROIs), where they are held until requested to be used by the Level 2 Trigger.

Level 2

The Level 2 trigger is software-based and is designed to reduce the triggering rate to 3.5 kHz. The typical process time is ≈ 10 ms. It uses full detector granularity and is seeded by level 1 ROIs, retrieved from the readout-buffers. At this stage information from the inner detector may also be used to further help in the identification of particles. By this procedure each ROI is subjected to a set of selection requirements called a hypothesis. ROIs not passing this are discarded.

Event Filter

The Event Filter is designed to select events which will be written to disk, and as such it reduces the trigger rate to give a maximum value of 200 Hz. It is controlled using offline algorithms/techniques and is able to employ the full detector granularity using fully reconstructed events. The event building process is performed by the Event Builder (EB) using the information from level 2 and the Read Out Buffer input cards (ROBins, which contain the data held in the pipeline memories after the level 1 selection was passed). Once the event is built it is passed to the Event Filter, where more sophisticated selections are applied to refine the Level 2 trigger objects. Events passing the event filter selections are directed into streams according to which trigger has been passed. Events passing more than one trigger that are associated with different streams will be written out to both.

Chapter 3

Signal and Background processes

This chapter explores the characteristics of the signal and discusses the main backgrounds to its search. The Monte Carlo simulated samples used are described together with the data used to explore whether there is good understanding within the Monte Carlo simulation.

3.1 Signal processes

The signal channel investigated in this thesis is a SM Higgs boson decaying to two Z bosons. As was shown in Section 1.4.2 this decay mode is one of the dominant ones over a large range of high Higgs masses. The decay of the Z bosons considered is with one Z decaying to leptons [1] and the other to neutrinos. Where one of the Z bosons decays to electrons(muons) this will subsequently be referred to as the electron(muon) channel (or $Z \to ee(Z \to \mu\mu)$ channel). An individual channel is not considered for the case of the τ lepton, but since τ decays typically involve electrons/muons, in this sense they are included. From Table 1.2, which gives a breakdown and branching ratios of the main Z decay modes, the main decays in this Higgs channel that include one Z decaying to leptons for triggering purposes are Z decays to leptons, leptons+jets and leptons+neutrinos. The decay to four leptons provides a signal which is the most easily identified in the detector, due to the presence of four high p_{T} leptons. However, it has the lowest BR, less than 0.1 %. In contrast the decay to leptons+jets has a much larger BR, $\approx 14\,\%$, but the presence of two jets makes it less easy to identify in the busy hadronic environment within ATLAS during data taking. The lepton+neutrino final state is perhaps a compromise between these. It has a BR in between the purely leptonic final state and the leptons+jets final state of $\approx 4\,\%$.

The lepton+neutrino final state has missing energy coming from the two neutrinos in the final state, which provides a good way to discriminate over background.

[1] Here, and subsequently, lepton refers to either an electron or a muon

This was demonstrated in recent studies performed in this channel using the GF production mode [40, 41], which showed that particularly at high mass where, because of the Higgs decay products becoming more boosted with increasing mass, the large missing energy present in the signal gives good sensitivity. At lower Higgs mass however, the signal has much lower missing energy making discrimination against background more difficult.

This study takes a first dedicated look at the VBF production mode, in particular to see if the characteristics associated with the VBF topology in this decay channel may provide an improvement in sensitivity. However, focussing on the VBF production mode is also motivated as within the SM it is predicted to provide a different mechanism by which a Higgs type boson could be produced, compared to the cross section dominant GF mode, and so must be studied in order to verify if this prediction is correct. Further, the VBF production mechanism provides access to different couplings compared to the GF mode [42], which will need to be studied in order to cross check our understanding of the mechanism through which the weak-gauge boson and fermion masses are generated.

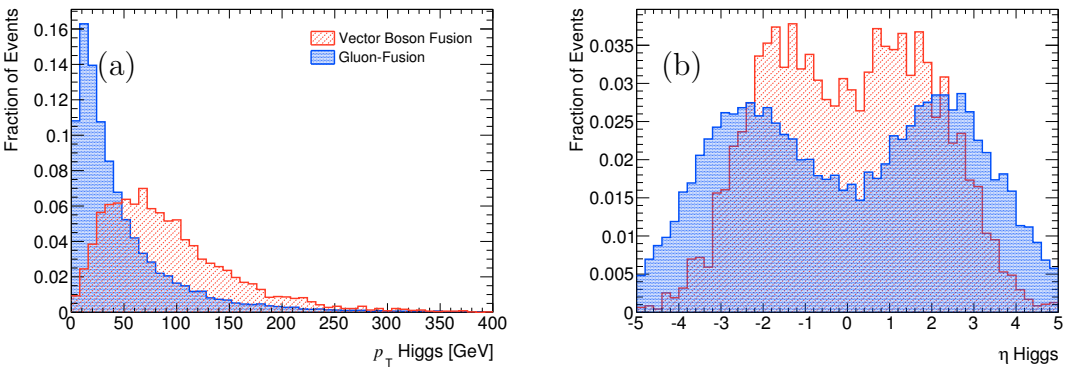


Figure 3.1: Comparison of Higgs boson (a) p_{T} and(b) η in vector boson fusion and gluon fusion produced truth $\mathrm{H} \rightarrow ll\nu\nu$ events ($\mathrm{m_H} = 200\,\mathrm{GeV}/c^2$)

Typically the p_{T} of a Higgs boson produced in a VBF event is harder than that in a GF event, as can be seen in Figure 3.1, which uses Monte Carlo *truth* level information in $\mathrm{H} \rightarrow \mathrm{ZZ} \rightarrow ll\nu\nu$ events. In addition it is more commonly in the central regions of the detector. These properties are a consequence of the Higgs boson recoiling off the typically forward *tag-jets* associated with the remnants of the VBF process (corresponding to the outgoing quarks produced in association with the Higgs boson as shown in Figure 1.10). Thus, the experimental signature of the signal in the VBF production mode is two hard isolated leptons, missing energy and tag-jets produced from the remnants of the VBF process.

3.2 Background processes

The signal is expected to suffer from various SM background processes, which can be categorised into the following

- top production via SM interactions
- Single and di-boson production (with heavy quarks or jets)
- QCD dijet and heavy quark production

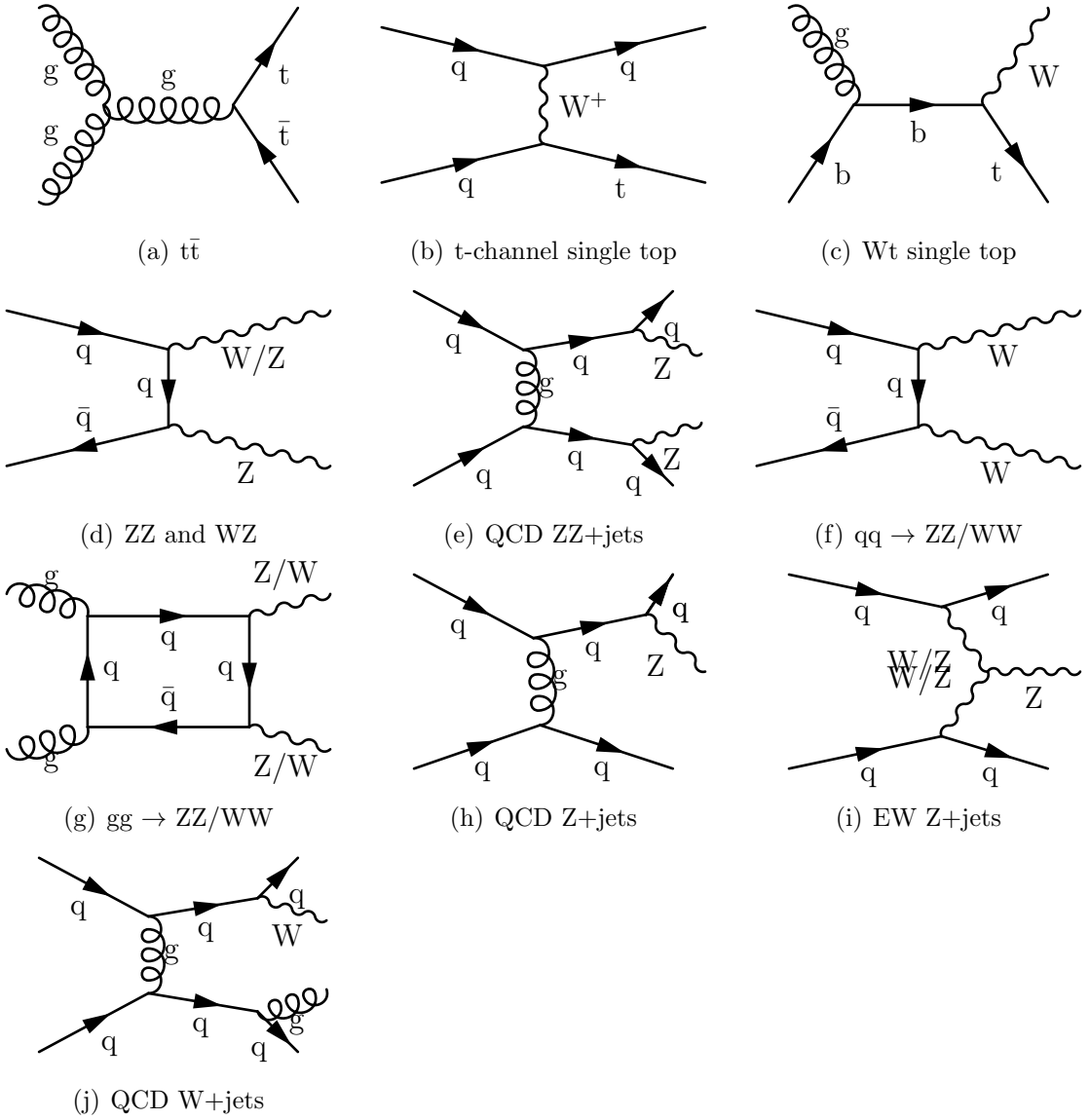
The production cross section for each of these backgrounds is orders of magnitude larger than the signal, requiring large background rejection in order for the signal to be detectable. Feynman diagrams of some of the backgrounds to this channel are shown in Figure 3.2. How each background can mimic the signal is discussed in the following sections.

Top pair production ($t\bar{t}$) The main decay modes of $t\bar{t}$ (Figure 3.2(a)) expected to contribute to the background to this analysis are where both of the Ws decay leptonically (lepton-lepton (ll) channel) and where one W decays leptonically and the other hadronically (lepton-hadron (lh) channel), making up $\approx 6(34)\,\%$ of all $t\bar{t}$ decays respectively. The lepton-lepton channel final state contains two leptons, missing energy and two b-quark jets. This, as shown below, becomes a background when the detector fails to identify both of the b-quark jets either because they lie outside the acceptance of the tracker or because these jets don't pass the b-tagging criteria.

$$\begin{array}{l} t\bar{t}\text{ (ll channel): } gg \to t\bar{t} \to WWb\bar{b} \to l\nu + \text{b}-\text{jet} + l\nu + \text{b}-\text{jet} \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \hookrightarrow \qquad\quad \hookrightarrow \text{not}-\text{tagged} \end{array}$$

The lepton-hadron channel becomes a background when one of the jets is misidentified as a lepton and the b-quark jets are not b-tagged (as indicated below). Because the top quarks recoil against each other when they are produced, the characteristics of the tag-jets in the signal may not provide much suppression of this background.

$$\begin{array}{l} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \curvearrowright \text{misidentified lepton} \\ t\bar{t}\text{ (lh channel): } gg \to t\bar{t} \to WWb\bar{b} \to \text{jets} + \text{b}-\text{jet} + l\nu + \text{b}-\text{jet} \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \hookrightarrow \qquad\quad \hookrightarrow \text{not}-\text{tagged} \end{array}$$



(a) $t\bar{t}$ (b) t-channel single top (c) Wt single top

(d) ZZ and WZ (e) QCD ZZ+jets (f) $qq \rightarrow ZZ/WW$

(g) $gg \rightarrow ZZ/WW$ (h) QCD Z+jets (i) EW Z+jets

(j) QCD W+jets

Figure 3.2: Feynman Diagrams of main backgrounds

Single top At the LHC the single top process has a much reduced cross section compared to $t\bar{t}$, but nonetheless should be considered as a background as it has similar final states. The dominant method of electroweak single top production at the LHC is via the t-channel exchange with a W boson (Figure 3.2(b)), making up approximately 80 % of the total single top production. However, in order to mimic the signal it must produce a fake lepton and forward jets which are not b-tagged. The Wt channel (Figure 3.2(c)) has a cross section of just less than 20 % of the total single top cross section and involves a top quark produced in combination with a W boson. It is likely to form a large component of the single top background due to the presence of the same final state as fully leptonic $t\bar{t}$ with one fewer b-quark jet. To mimic the signal, additional jets from initial/final state radiation, underlying event or pile-up misidentified as tag-jets must be produced. How each component can fake the signal is shown below

$$\begin{aligned} \text{t-channel}: \quad & q\bar{q} \rightarrow tq \rightarrow Wbq \rightarrow l\nu + \text{b-jet} + \text{jet} \\ & \hookrightarrow \text{not-tagged} \\ & \hookrightarrow + \text{additional lepton} \end{aligned}$$

$$\begin{aligned} \text{Wt channel}: \quad & q\bar{q} \rightarrow Wt \rightarrow l\nu + l\nu + \text{b-jet} \\ & \hookrightarrow \text{not-tagged} \\ & \hookrightarrow + \text{additional jets} \end{aligned}$$

The s-channel has a much smaller cross section compared to either the t-channel or Wt production mechanisms and corresponds to a few percent of the total single top cross section. It has a similar final state to the t-channel, but whereas in the t-channel the quark which is produced in association with the top quark is not derived from b hadrons, in the s-channel process it is. Therefore the same requirements as those of the t-channel need to be met if this process is to mimic the signal, but in addition the b-quark jet produced in association with the top must not be identified as such. Given this and its small contribution to the total single top cross section, the s-channel is not considered in this analysis.

ZZ, WW and WZ Although the di-boson backgrounds (WW, WZ and ZZ), including production with an additional number of jets in the final state (i.e. Figures 3.2(d)-3.2(g)), have cross sections typically many orders of magnitude smaller than $t\bar{t}$, they can often involve final states very similar to that found in the signal topology, in particular the two high p_{T} leptons. In ZZ these leptons will be derived from a Z boson as in the signal. When the other Z decays to neutrinos, the tag-jets in the signal may then be faked from the presence of additional jets in the busy hadronic

environment at the LHC. In cases where this Z decays to jets, a mis-measurement of missing energy is required for the signal to be faked.

WW can also form a final state with similar characteristics as the decay products of the Higgs signal. This occurs when each W decays to a lepton plus neutrino. As with ZZ, additional jets may lead to a final state the same as that in the signal. However, the fact that the leptons in WW come from the decay of W bosons as opposed to Z bosons will allow a certain level of discrimination of such cases.

WZ can involve Z decays to leptons and give rise to large amounts of missing energy through the W decay to lepton and neutrino. In this case, to suppress this background it will be important to veto events which have any other leptons other than the two expected in the signal.

$$\begin{aligned} \mathrm{q\bar{q}} \rightarrow \mathrm{ZZ} \rightarrow ll + \nu\nu & \\ & \hookrightarrow + \text{additional jets} \\ \mathrm{q\bar{q}} \rightarrow \mathrm{WW} \rightarrow l\nu + l\nu & \\ & \hookrightarrow + \text{additional jets} \\ \mathrm{q\bar{q}} \rightarrow \mathrm{WZ} \rightarrow l\nu + ll & \\ & \hookrightarrow \text{mis-id jet} + \text{additional jet} \end{aligned}$$

Z+jets Z produced in association with jets (typically light quark jets) (Figures 3.2(h)-3.2(i)) can lead to events with two high p_{T} leptons from the Z decay and additional jets which might mimic the tag-jets found in the signal. In this case Z+jets becomes a background if a mis-measurement of missing energy leads to a non-negligible level in the final state. In this study the decays of Z's to electrons, muons and taus are considered. In the case of Z$\rightarrow \tau\tau$ the same decay products as in the signal will occur through decays of the taus to, for example, leptons and neutrinos as shown below. Production with an additional number of jets ranging from zero to five is considered. The contribution from Z produced in association with a $\mathrm{b\bar{b}}$ pair is also considered. It becomes a background when, in addition to there being a mis-measurement of missing energy, the two b-quark jets are not b-tagged. This extra requirement in order to mimic the signal suggests this process will have a small contribution to the total background from Z+jets.

$$\left\{ \begin{array}{l} e : \mathrm{q\bar{q}} \rightarrow Z + \text{jets} \rightarrow ee + \text{jets} \\ \mu : \mathrm{q\bar{q}} \rightarrow Z + \text{jets} \rightarrow \mu\mu + \text{jets} \end{array} \right\} + \text{mis-measured missing energy}$$

$$\tau\text{: } \mathrm{q\bar{q}} \rightarrow \text{Z+jets} \rightarrow \tau\tau + \text{jets} \rightarrow l\nu\nu + l\nu\nu + \text{jets}$$

W+jets W+jets (Figure 3.2(j)) constitutes a background if an additional jet is identified as a lepton. Although this is expected to happen rarely, its cross section can be over an order of magnitude larger than that for Z+jets production and as such it must be taken into account. Due to the presence of two leptons in the signal, W→,$e\nu, \mu\nu$,$\tau\nu$ decays are considered. In this analysis three different production channels are used, W, Wbb and Wc, where in each case the W is produced in association with a number of jets, ranging upward from zero. Wbb,c become a background when one of the heavy quarks decays semi-leptonically, producing the additional lepton. For example b quarks decay via b→$cl\nu$ approximately 20 % of the time. For these cases to yield an energetic lepton the parent b/c meson must also be very energetic, meaning its decay products are collimated. This implies that the energetic leptons produced in such processes will not be isolated in space from other particles as expected in the signal and as such these processes will only be a background when the lepton appears to be isolated.

QCD The cross section for QCD processes including production of heavy quark pairs and QCD di-jets is many orders of magnitude greater than any of the other backgrounds considered. Therefore, even if a very small fraction of these events pass the selection criteria, the resulting background would likely be quite large. However in each case to be a background these types of processes would need to have jets mimic the forward nature of the tag-jets in the signal and have a non-negligible level of missing energy. In addition to this, the two energetic and isolated leptons of the signal must be found. Heavy quark decays would form such leptons if as detailed earlier in this section, the mesons with which the quarks are associated, decayed to leptons, shown below. However, leptons produced in this way will rarely have the same properties as the leptons in the signal, i.e. be isolated, thereby indicating the contribution of this background can be highly diminished.

$$\mathrm{gg} + \mathrm{q\bar{q}} \rightarrow \mathrm{b\bar{b}} \rightarrow cl\nu + cl\nu$$

↳ leptons measured as isolated

↳ + additional jets + missing energy

QCD di-jets would form a background if the jets were able to fake the leptons in the signal with a high enough rate. However, ATLAS is designed to be able to detect leptons with an efficiency of over 70 % and be able to correctly distinguish between jets and leptons such that only 1 in approximately 10000 jets is wrongly identified as a lepton. In this way the background from QCD di-jets can also be largely suppressed (see Section 5.4).

3.3 Phenomenology at hadron colliders and Monte Carlo simulation

The description of the Higgs signal and the explanation of its background processes earlier in this chapter do not take into account the fact that the LHC collides composite protons (A, B). In order to model the interaction of composite protons at the energies produced by the LHC it is useful to consider the parton model in which the proton is made up of constituent partons (a, b). In this model the interaction of individual partons leads to the production of other particles such as the Higgs boson (c), the production cross section of which, $d\sigma_{a+b\to c}$, can be found using the SM. In addition other remnants (Z) will be produced. The calculation of the hadronic cross section $(d\sigma_{A+B\to c+Z})$ however, must be calculated within the parton model according to

$$d\sigma_{A+B\to c+Z} = \sum_{a,b} \int_0^1 dx_a \int_0^1 dx_b f_A^a(x_a, Q^2) f_B^b(x_b, Q^2) d\sigma_{a+b\to c} \tag{3.1}$$

where the sum is over all processes (i.e. Feynman diagrams) contributing to the production of c. f_A^a and f_B^b are called *parton distribution functions* (PDFs) and correspond to the probability to have a parton a with momentum fraction x_a within its parent proton A at energy scale Q. These cannot be calculated from first principles but must be measured for example in deep inelastic scattering experiments. In this way the total hadronic cross section is composed of two parts, a perturbative short distance scale part and a non-perturbative long distance scale part. The procedure of separating the interaction like this is called factorization and the energy scale at which this is done is called the factorization scale.

A diagram showing the different processes occurring when high energy protons collide is shown in Figure 3.3. The interaction of the protons leading to, for example, the production of the Higgs boson represented by $d\sigma_{A+B\to c+Z}$ is calculated within the parton model and corresponds to the hard sub-process. All other contributions to the final state not originating from the hard sub-process are called the underlying event. This includes initial state radiation (ISR) produced via emission from the incoming partons, interactions between the proton remnants (i.e. partons other than those in the hard interaction) and any final state radiation (FSR) from the final state particles. Because the final state partons carry colour charge they often radiate gluons, leading to production of quark anti-quark pairs. This gives rise to cascades of partons called parton showers. Once energetically favourable, the partons produced in such showers form colour neutral states in the process of hadronization. The

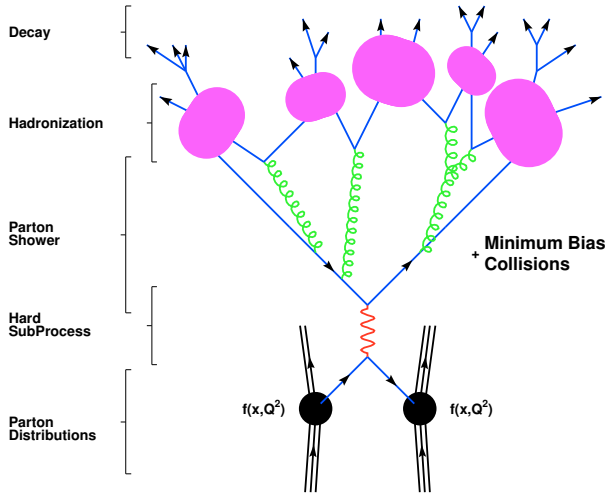


Figure 3.3: Phenomenological model of the interaction of a proton-proton collision at high energy scale. [43]

decay products of these states are then subsequently measured in the detector. Understanding of parton showering and hadronization is achieved using dedicated models.

The most common technique to model physical processes is to use Monte Carlo methods. Monte Carlo methods use pseudo-random numbers to model particle interactions based on the underlying physical principles outlined in Chapter 1. In particle physics this procedure is generally referred to as event generation and the program using the Monte Carlo methods is called the Monte Carlo generator. It allows the kinematics of any final state particles to be calculated given an input process and a set of initial starting conditions. The final state particles correspond to those that are stable in the sense that the distance they travel in the associated particle's proper lifetime is within a suitably large range. Some Monte Carlo generators simulate specific final states (called matrix element generators). This is typically done by summing over all relevant Feynman diagrams. An example is MC@NLO [44]. Other generators simulate non perturbative effects including hadronisation. An example is PYTHIA [45]. Details of the particles produced in the event generation are provided in the Monte Carlo *truth* information.

All Monte Carlo officially produced for the ATLAS Collaboration is produced within the ATHENA framework. This is a software framework which provides interfaces to all the generators used. In order to get a realistic picture of what we would expect to see in the detector when a particular process occurs, the interactions of the final state particles and the detector material are modelled. This procedure is done with the GEANT program and also performed in the ATHENA framework.

Two approaches to this exist and are referred to as fast/full simulation respectively, depending on the level of detail at which the detector simulation is done.

3.4 Monte Carlo samples

The Monte Carlo event samples used in this study were officially produced within the ATLAS Collaboration, using version 15.6.3 of the ATHENA framework. They were fully simulated using GEANT 4 and correspond to p-p collisions at $\sqrt{s} = 7\,\text{TeV}$. The default samples used in this analysis do not model pile-up and are used to obtain the main results. This choice was made because of insufficient statistics in some of the pile-up samples for the backgrounds considered and because not all the processes considered were simulated taking pile-up into account. However, where possible the effect of pile-up was investigated. Where comparisons to pile-up Monte Carlo are made, this is explicitly stated and the pile-up Monte Carlo is referred to as such. The term Monte Carlo used without mention of pile-up corresponds to the non pile-up Monte Carlo. The pile-up samples considered were simulated with two interactions per bunch crossing. In the following the main properties of the simulated samples used will be discussed.

3.4.1 Signal

The signal samples used were generated with PYTHIA 6.421 [45]. Z decays to leptons included electrons, muons and taus in their corresponding branching fractions. ISR was modelled with PHOTOS [46] and the decay of tau leptons by TAUOLA [47]. The events include both VBF and GF contributions. The cross sections used are from [19], with GF quoted at NNLO and VBF at NLO. The analysis uses signal samples simulated for Higgs masses between m_H = 200-600 GeV/c^2 in 20 GeV/c^2 steps. The main properties of the signal samples used, including number of simulated events, cross section and the corresponding integrated luminosity are shown in Table 3.1. A break-down of the VBF and GF components is shown. Separation of the samples into VBF and GF components was done using truth information. This was done by identifying the Higgs boson and then navigating backwards to identify if it was derived from gluons or quarks and so produced by GF or VBF. The identity of each particle was found using the PDG particle codes [1]. In order to verify that the separation of VBF and GF events was done correctly the fraction of VBF events obtained was compared to that produced in a statistically independent sample of $H \to ZZ \to llνν$ produced with the same configuration options as the official ATLAS Monte Carlo. Agreement was found to within 1 % for the range of mass samples

Higgs Mass	N_{MC}			σ [fb]			Integrated Luminosity [fb^{-1}]		
	VBF	GF	Total	VBF	GF	Total	VBF	GF	Total
				Non-Pile-up					
200	6575	23415	29990	6.59	54.28	60.87	997.95	431.36	492.69
220	6711	23232	29943	6.22	49.19	55.41	1079.27	472.25	540.37
240	6835	23152	29987	5.51	42.53	48.04	1240.06	544.38	624.19
280	6919	23074	29993	4.82	37.10	41.92	1436.41	621.86	715.45
300	6936	23005	29941	4.25	33.11	37.36	1631.91	694.85	801.46
320	6817	23173	29990	3.73	29.99	33.72	1826.21	772.77	889.39
360	6477	23467	29944	3.27	28.06	31.33	1978.99	836.30	955.66
380	5906	24076	29982	2.87	27.63	30.50	2056.45	871.49	983.07
400	5258	24736	29994	2.42	28.30	30.72	2171.72	873.99	976.26
440	4940	24051	28991	2.04	25.51	27.54	2425.64	942.99	1052.62
460	4915	25073	29988	1.76	22.11	23.87	2793.74	1133.84	1256.16
480	5125	24867	29992	1.54	18.51	20.06	3317.69	1343.09	1495.23
520	5273	24723	29996	1.37	15.52	16.89	3850.07	1592.63	1775.65
540	5658	24335	29993	1.23	12.97	14.20	4614.21	1875.58	2112.06
560	5964	24024	29988	1.10	10.78	11.88	5403.15	2228.66	2523.52
600	6373	23618	29991	1.00	8.96	9.96	6369.40	2636.32	3011.37
				Pile-up					
200	6575	23415	29990	6.59	54.28	60.87	997.95	431.36	492.69
300	6936	23005	29941	4.25	33.11	37.36	1631.91	694.85	801.46
400	5258	24736	29994	2.42	28.30	30.72	2171.72	873.99	976.26
600	6373	23618	29991	1.00	8.96	9.96	6369.40	2636.32	3011.37

Table 3.1: Summary of signal Monte Carlo sample properties as a function of Higgs mass used in this analysis, including number of simulated events (N_{MC}), cross section (σ [fb]) and corresponding integrated luminosity [fb^{-1}]. Included are the relative contributions from VBF and GF production mechanisms. The samples were generated with PYTHIA. Pile-up samples were simulated with two interactions per bunch crossing.

tested (m_H = 200-600 GeV/c^2 in 100 GeV/c^2 steps) indicating that the separation of VBF and GF was performed correctly.

A comparison between VBF and GF cross sections at $\sqrt{s} = 7$ TeV with varying Higgs mass is shown in Figure 3.4. The cross section is reduced by a factor six with increasing mass in the mass range investigated.

PYTHIA-HERWIG VBF production comparison The HERWIG generator is widely used to simulate VBF processes. However, in this study, the available signal samples were generated with PYTHIA. A comparison of some basic properties of the truth level partons associated with the VBF process (VBF quarks) and partons other than this produced in the final state with PYTHIA and HERWIG was made in order to investigate whether the differences in the generators produce any large discrepancies. The PYTHIA sample used was composed of VBF and GF events and the VBF component was identified using the same method discussed in the last section. The HERWIG sample consisted purely of VBF events. Both samples used were of a $m_H = 120\,GeV/c^2$ Higgs boson decaying to photons as these were the only samples available for which a direct comparison between the generators could be

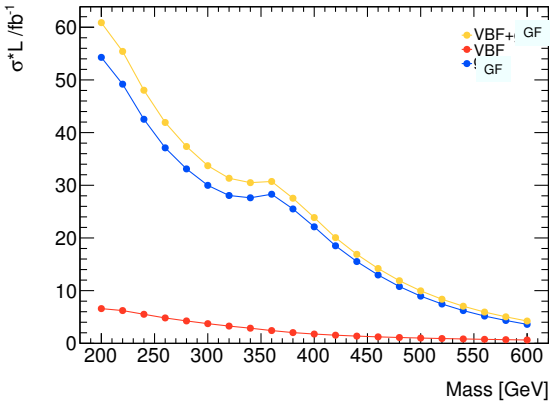


Figure 3.4: Comparison of predicted number of VBF and GF events produced at $\sqrt{s} = 7\,\text{TeV}$ for $1\,\text{fb}^{-1}$ as a function of Higgs mass.

made.

The p_{T} distribution of the truth level VBF quarks identified in the respective PYTHIA and HERWIG samples, are shown in Figure 3.5(a). Both distributions are in agreement over the p_{T} range considered to within the level of 1%. The same agreement is found for the η and ϕ distributions, shown in Figures 3.5(b) and 3.5(c) respectively.

In Figure 3.6, the generator level comparison for the (a) p_{T}, (b) η and (c) ϕ distributions of additional quarks other than those connected with the VBF topology is shown. Here the agreement between the generators is less good, with the additional quarks having a harder p_{T} scale and being more forward in PYTHIA compared to HERWIG. However the distributions are still within agreement at the level of a few percent. The discrepancy in η was found to be largely due to low p_{T} candidates. It is perhaps understandable that the distributions of the additional quarks show a poorer agreement because they depend on the global kinematics of the whole event. Historically there was a preference within ATLAS to use HERWIG for VBF related studies. However, the recently produced samples were made using PYTHIA, which are therefore used for practical reasons.

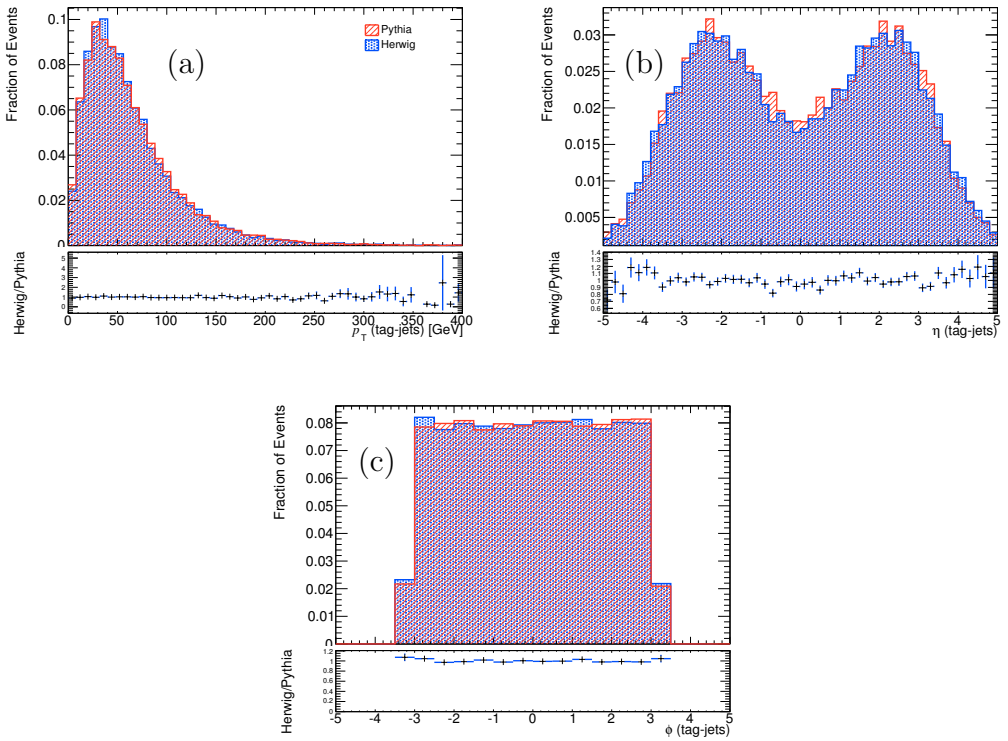


Figure 3.5: Comparison of PYTHIA and HERWIG generators truth Monte Carlo quarks associated with the VBF process for (a) p_{T} (b) η and (c) ϕ distributions (using $\mathrm{m_H} = 120\,\mathrm{GeV}/c^2$ H$\rightarrow \gamma\gamma$ samples)

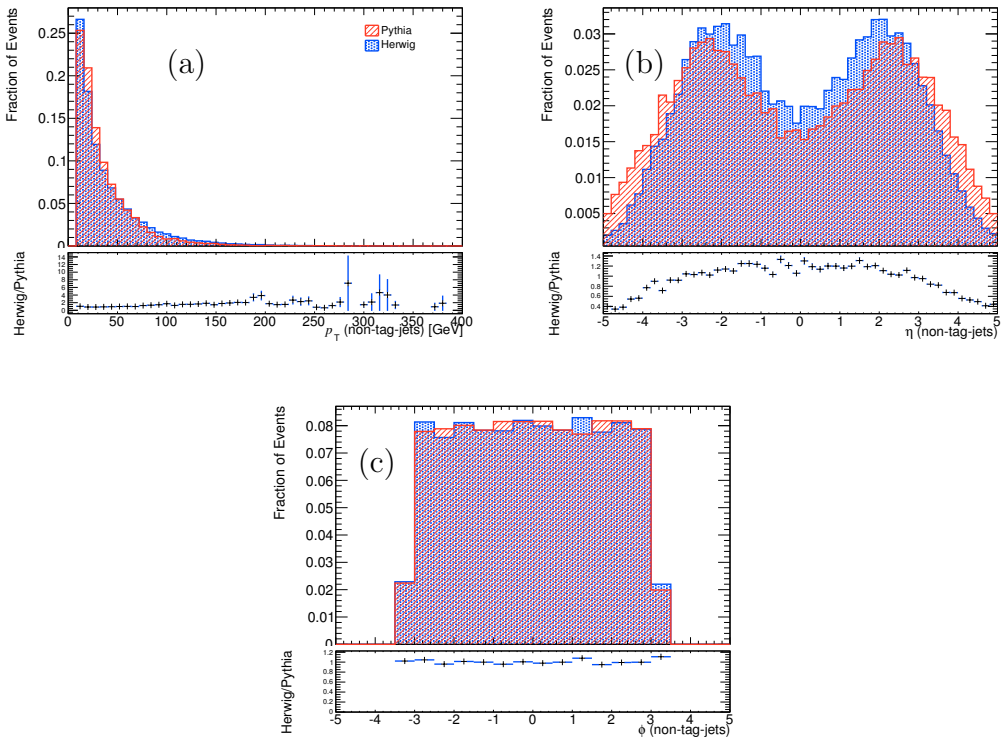


Figure 3.6: Comparison of PYTHIA and HERWIG generators for truth additional-quarks (a) p_{T} (b) η and (c) ϕ distributions (using $120\,\mathrm{GeV}/c^2$ H$\rightarrow \gamma\gamma$ samples)

Process	N_{MC}		σ [fb]	Integrated Luminosity [fb^{-1}]	
	non pile-up	pile-up		non pile-up	pile-up
$t\bar{t}$	999387	999387	91550.52	10.92	10.92
single top (t channel $e\nu$)	9993	9993	2317.51	4.31	4.31
single top (t channel $\mu\nu$)	9997	9997	2325.02	4.30	4.30
single top (t channel $\tau\nu$)	10000	10000	2309.47	4.33	4.33
single top (Wt)	14995	56969	6386.48	2.35	8.92

Table 3.2: Summary of top Monte Carlo sample properties used in this analysis, including number of simulated events (N_{MC}), cross section (σ [fb]) and corresponding integrated luminosity [fb^{-1}]. The samples were generated with MC@NLO. For the $t\bar{t}$ sample a filter is applied in which at least one W is required to decay leptonically. The cross section quoted for the $t\bar{t}$ sample is NLO accuracy and taken from [49] and then multiplied by a filter efficiency of 0.5562 [40]. The single top cross section used are from the generator.

3.4.2 Backgrounds

The following section outlines details of the background Monte Carlo samples used.

Top background Both contributions from $t\bar{t}$ and single top were simulated with MC@NLO [44] interfaced with JIMMY v4.1 [48]. The main characteristics of the top samples used are shown in Table 3.2. The $t\bar{t}$ sample is filtered at generator level to require at least one lepton (e, μ, τ) $p_T > 1\,\mathrm{GeV}/c$ originating from a W boson (i.e. only events with at least one leptonic W retained). The $t\bar{t}$ cross section used is 164.6 pb [49], which is based on calculations performed in [50, 51] using the HATHOR program [52] using the PDF set CTEQ66 [53] and assuming a top quark mass of $172.5\,\mathrm{GeV}/c^2$. This is then combined with a filter efficiency of 0.5562 [40]. t-channel single top sample decays to e, μ and τ are considered. The single top cross sections were taken from the generator prediction (Wt channel 66.2 pb and t-channel 14.6 pb) and then convoluted with branching ratios from [1] relevant to each decay type.

Di-boson backgrounds All nominal di-boson samples used were simulated with MC@NLO interfaced with JIMMY 4.31 for simulation of the underlying event. A summary of the properties of each sample considered is given in Table 3.3. The cross sections listed correspond to the cross sections quoted in the following including the branching fractions from [1]. In the case of ZZ, individual decays to llqq, $4l$, $ll\nu\nu$, 4τ, $ll\tau\tau$, $\tau\tau\nu\nu$ and $\tau\tau$qq (where $l = e$ or μ) are considered. The ZZ cross section used is based on the NLO prediction [54]. A 6 % correction (suggested in Ref [55]) to account for missing gluon-pair quark box diagrams is made to this value such that the ZZ cross section used is 5.96 pb.

For the WW samples, decays to lepton and neutrino are considered, where the lepton can be e, μ or τ. The WW cross section used is 44.9±2.2 pb [56, 57]. It is expected WW production through gluon fusion will contribute $\approx 3\,\%$ to overall WW cross section. This is taken into account by scaling the total cross section by an additional 3 %.

For the WZ background decays to $l\nu ll, l\nu$qq, llqq, $l\nu\tau\tau$, $\tau\nu ll$ and $\tau\nu\tau\tau$ are considered. The cross section used is 18.0 pb [54].

In addition to the above nominal di-boson samples, a different set of di-boson samples produced with between 0-3 additional partons in the final state are used in order to estimate the contribution of di-boson+jets. However, because only a small number of events and a limited number of boson decays are generated in these samples, they are used as a systematic. Samples used include WW decaying to $l\nu l\nu$, WZ to inclll and ZZ to inclll, where incl means the inclusive decay of the corresponding boson. These samples were generated with ALPGEN interfaced with HERWIG. Cross sections were taken from the generator prediction and include a k-factor of 1.21 [58].

Process	N_{MC}	σ [fb]	Integrated Luminosity [fb^{-1}]
	non pile-up/pile-up		non pile-up/pile-up
WW			
$\rightarrow e\nu e\nu$	49990	534.25	93.57
$\rightarrow e\nu\mu\nu$	49944	525.30	95.08
$\rightarrow e\nu\tau\nu$	49944	559.09	89.33
$\rightarrow \mu\nu\mu\nu$	49939	516.50	96.69
$\rightarrow \mu\nu e\nu$	49988	525.30	95.16
$\rightarrow \mu\nu\tau\nu$	49942	549.73	90.85
$\rightarrow \tau\nu\tau\nu$	49937	585.10	85.35
$\rightarrow \tau\nu e\nu$	49942	559.09	89.33
$\rightarrow \tau\nu\mu\nu$	49992	549.73	90.94
ZZ			
$\rightarrow ll$qq	24994	591.00	42.29
$\rightarrow 4l$	99986	24.68	4051.30
$\rightarrow ll\nu\nu$	99840	160.42	622.37
$\rightarrow 4\tau$	24995	6.77	3692.73
$\rightarrow 2l2\tau$	24990	27.03	924.51
$\rightarrow \tau\tau\nu\nu$	24994	80.34	311.10
$\rightarrow \tau\tau$qq	24993	280.80	89.01
W^+Z			
$\rightarrow l\nu ll$	24997	167.15	149.55
$\rightarrow \tau\nu ll$	24996	87.06	287.12
$\rightarrow l\nu\tau\tau$	24995	83.71	298.59
$\rightarrow \tau\nu\tau\tau$	24998	43.60	573.36
$\rightarrow ll$qq	24987	521.80	47.89
$\rightarrow l\nu$qq	24997	1674.96	14.92
W^-Z			
$\rightarrow l\nu ll$	99884	94.48	1057.25
$\rightarrow \tau\nu ll$	24996	49.21	507.99
$\rightarrow l\nu\tau\tau$	24997	47.31	528.31
$\rightarrow \tau\nu\tau\tau$	24996	24.64	1014.32
$\rightarrow ll$qq	99922	294.93	338.80
$\rightarrow l\nu$qq	24989	946.72	26.40
WW $\rightarrow l\nu l\nu$			
+ 0 jets	49842	2479.29	20.10
+ 1 jets	24995	1194.27	20.93
+ 2 jets	14996	533.61	28.10
+ 3 jets	9997	215.38	46.42
WZ $\rightarrow$ inclll			
+ 0 jets	14999	803.44	18.67
+ 1 jets	9996	482.79	20.70
+ 2 jets	4995	267.41	18.68
+ 3 jets	4997	112.53	44.41
ZZ $\rightarrow$ inclll			
+ 0 jets	9997	597.74	16.72
+ 1 jets	4500	272.25	16.53
+ 2 jets	4993	106.48	46.89
+ 3 jets	2498	33.88	73.73

Table 3.3: Summary of di-boson Monte Carlo sample properties used in this analysis, including number of simulated events (N_{MC}), cross section (σ [fb]) and corresponding integrated luminosity [fb^{-1}]. The samples shown in the upper part of the table were generated with MC@NLO interfaced with JIMMY. The samples shown in the bottom part of the table were generated with ALPGEN interfaced with Herwig. Incl refers to the inclusive decay of one of corresponding W or Z bosons. l refers to either e or μ.

Z+jets Z+jets samples were simulated with ALPGEN v2.13 [59] interfaced with HERWIG v6.510 [60] for modelling of parton showers and hadronization. The simulated samples include generated hard matrix elements for Z and Zb$\bar{\text{b}}$ with additional numbers of partons in the final state: 0-5 partons for the Z samples and 0-3 partons for the Zb$\bar{\text{b}}$ samples. In each case Z decays to $ee, \mu\mu$ and $\tau\tau$ are considered. The cross sections were taken from the generator prediction. A k-factor of 1.22 [29] is used to scale the LO generator cross section to NLO. Details of the Z+jets samples used are shown in Table 3.4.

Process	N_{MC}		σ [fb]	Integrated Luminosity [fb^{-1}]	
	non pile-up	pile-up		non pile-up	pile-up
$Z \to ee$					
+ 0 jets	6614832	303966	804956	8.22	0.38
+ 1 jets	1333815	62941	162321	8.22	0.39
+ 2 jets	404755	18997	48739	8.30	0.39
+ 3 jets	109954	5499	13603	8.08	0.40
+ 4 jets	29982	1499	3355	8.94	0.45
+ 5 jets	9989	500	976	10.23	0.51
$Z \to \mu\mu$					
+ 0 jets	6610859	303947	804956	8.21	0.38
+ 1 jets	1334502	62996	162321	8.22	0.39
+ 2 jets	404883	18993	48739	8.31	0.39
+ 3 jets	109964	5497	13603	8.08	0.40
+ 4 jets	29981	1499	3355	8.94	0.45
+ 5 jets	9994	499	976	10.24	0.51
$Z \to \tau\tau$					
+ 0 jets	6615633	302959	802028	8.25	0.38
+ 1 jets	1334329	62981	162260	8.22	0.39
+ 2 jets	404679	18993	49288	8.21	0.39
+ 3 jets	109888	5497	13420	8.19	0.41
+ 4 jets	29985	1499	3538	8.48	0.42
+ 5 jets	9992	499	854	11.70	0.58
$Z(\to ee)b\bar{b}$					
+ 0 jets	149925	149925	7954.4	18.85	18.85
+ 1 jets	99973	99973	3013.4	33.18	33.18
+ 2 jets	39989	39989	985.76	40.57	40.57
+ 3 jets	9949	9949	472.14	21.07	21.07
$Z(\to \mu\mu)b\bar{b}$					
+ 0 jets	149968	149968	7954.4	18.85	18.85
+ 1 jets	99975	99975	3013.4	33.18	33.18
+ 2 jets	39988	39988	985.76	40.57	40.57
+ 3 jets	9997	9997	472.14	21.17	21.17
$Z(\to \tau\tau)b\bar{b}$					
+ 0 jets	149821	149821	7954.4	18.83	18.83
+ 1 jets	99921	99921	3013.4	33.16	33.16
+ 2 jets	39982	39982	985.76	40.56	40.56
+ 3 jets	9996	9996	472.14	21.17	21.17

Table 3.4: Summary of Z+jets Monte Carlo sample properties used in this analysis, including number of simulated events (N_{MC}), cross section (σ [fb]) and corresponding integrated luminosity [fb^{-1}]. The samples were generated with ALPGEN interfaced with HERWIG.

Process	N_{MC}		σ [fb]	Integrated Luminosity [fb^{-1}]	
	non pile-up	pile-up		non pile-up	pile-up
$W \to e\nu$					
+ 0 jets	1381931	1381531	4640000	0.30	0.30
+ 1 jets	258408	641645	1577460	0.16	0.41
+ 2 jets	188896	188896	460062	0.41	0.41
+ 3 jets	50477	49978	123098	0.41	0.41
+ 4 jets	12991	12991	30866	0.42	0.42
+ 5 jets	3449	3449	8418	0.41	0.41
$W \to \mu\nu$					
+ 0 jets	1386038	1386038	4640000	0.30	0.30
+ 1 jets	255909	255909	1563064	0.16	0.16
+ 2 jets	187860	187860	457866	0.41	0.41
+ 3 jets	50887	49887	123342	0.41	0.40
+ 4 jets	12991	12991	31354	0.41	0.41
+ 5 jets	3498	3498	8540	0.41	0.41
$W \to \tau\nu$					
+ 0 jets	1365491	1364841	4640000	0.29	0.29
+ 1 jets	254753	254753	1557696	0.16	0.16
+ 2 jets	188446	187946	459452	0.41	0.41
+ 3 jets	50472	49972	122976	0.41	0.41
+ 4 jets	12996	12996	31354	0.41	0.41
+ 5 jets	3998	3998	8540	0.47	0.47
$Wb\bar{b}$					
+ 0 jets	6499	6499	3904	1.66	1.66
+ 1 jets	5500	5500	3172	1.73	1.73
+ 2 jets	2997	2997	1708	1.75	1.75
+ 3 jets	1500	1500	732	2.05	2.05
Wc					
+ 0 jets	862565	862565	526186	1.64	1.64
+ 1 jets	320379	319929	195322	1.64	1.64
+ 2 jets	84899	84899	51850	1.64	1.64
+ 3 jets	19992	19992	12078	1.66	1.66
+ 4 jets	4995	4995	2806	1.78	1.78

Table 3.5: Summary of W+jets Monte Carlo sample properties used in this analysis, including number of simulated events (N_{MC}), cross section (σ [fb]) and corresponding integrated luminosity [fb^{-1}]. The samples were generated with ALPGEN interfaced with HERWIG.

W+jets W+jets samples were simulated with ALPGEN v2.13 [59] interfaced with HERWIG v6.510 [60] for modelling of parton showers and hadronization. The simulated samples include generated hard matrix elements for W,$Wb\bar{b}$ and Wc with additional numbers of partons in the final state, 0-5 for the W samples, 0-3 for the $Wb\bar{b}$, and 0-4 for the Wc samples. In each case W decays to $e, \mu, \tau\nu$ are considered. The cross sections were taken from the generator prediction. A k-factor of 1.20 [40] is used to scale the LO generator cross section to NLO. Details of the W+jets samples used are shown in Table 3.5.

Process QCD dijets	N_{MC}	σ [pb]	Integrated Luminosity [pb^{-1}]
J0 ($0< p_T <17$)	1399184	9.8568×10^{9}	1.42×10^{-4}
J1 ($17< p_T <35$)	1395383	6.7808×10^{8}	2.06×10^{-3}
J2 ($35< p_T <70$)	1398078	4.0994×10^{7}	3.41×10^{-2}
J3 ($70< p_T <140$)	1397430	2.1936×10^{6}	6.37×10^{-1}
J4 ($140< p_T <280$)	1397401	8.7704×10^{4}	1.59×10^{1}
J5 ($280< p_T <560$)	1391612	2.3498×10^{3}	5.92×10^{2}
J6 ($560< p_T <1120$)	1347654	3.3615×10^{1}	0.40×10^{5}
J7 ($1120< p_T <2240$)	1125428	1.3741×10^{-1}	8.19×10^{6}
J8 ($p_T >2240$)	1383585	6.2144×10^{-6}	2.22×10^{11}
$b\bar{b}$			
$\rightarrow ee$	4447997	7.39×10^{7}	6.02×10^{-2}
$\rightarrow \mu\mu$	4443898	7.39×10^{7}	6.01×10^{-2}
$c\bar{c}$			
$\rightarrow ee$	1494456	2.84×10^{7}	5.26×10^{-2}
$\rightarrow \mu\mu$	1499257	2.84×10^{7}	5.28×10^{-2}

Table 3.6: Summary of QCD Monte Carlo sample properties used in this analysis, including number of simulated events (N_{MC}), cross section (σ [pb]) and corresponding integrated luminosity [pb^{-1}]. The samples were generated with PYTHIA. p_T thresholds are in GeV.

QCD background The contribution to QCD samples from heavy quark decays is measured using $b\bar{b}$ and $c\bar{c}$. Decays to electrons and muons are considered in each case, requiring a Monte Carlo truth lepton from the quark decay with $p_T > 15\,\text{GeV}/c$ and $|\eta| < 2.5$ is present. These samples were generated with PYTHIA 6.421 [45] using the PYTHIAB interface. The QCD di-jet samples used were also simulated with PYTHIA 6.421. This contribution is divided into different sub-samples corresponding to the p_T threshold of the hard sub-process. For example, J1 corresponds to the threshold $17< p_T <35\,\text{GeV}/c$. For all QCD samples used the LO cross section was used [61]. Their details are shown in Table 3.6.

3.5 Data sample

The data used in this study is from the ATLAS 2010 dataset, in which the LHC was operated at $\sqrt{s} = 7\,\text{TeV}$ between 30^{th}Mar-29^{th}Oct 2010. During this time a total integrated luminosity of $48.8\,\text{pb}^{-1}$ was delivered to ATLAS, of which $46.72\,\text{pb}^{-1}$ was recorded. Figure 3.7 shows (a) the total integrated luminosity and (b) the peak luminosity delivered as a function of time in the 2010 data collection run. During the 2010 data taking period, the data taking sessions were divided up into periods, with each period made of a number of runs, each given a unique identifying number. A typical run involves a period of stable data taking whereby typically, beams are injected into the LHC, stable beams are declared and data taking is maintained until beams are dumped or lost. The data in 2010 is composed of periods A to I. Each run is composed of fixed luminosity intervals called lumi-blocks. Within ATLAS,

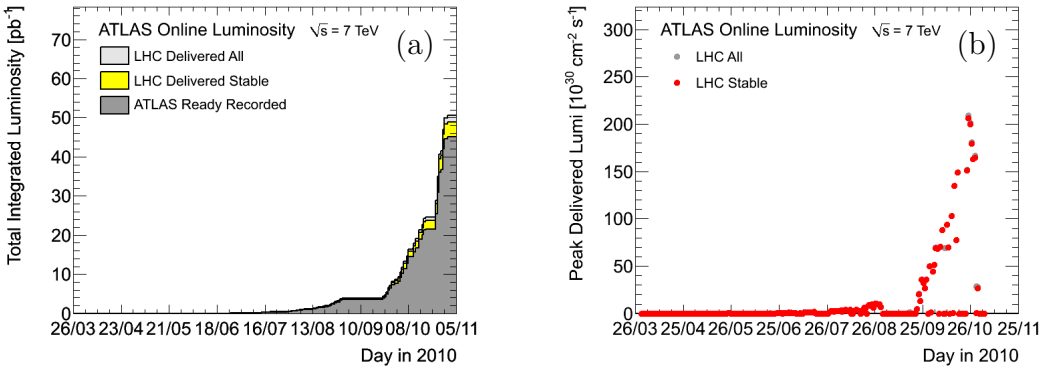


Figure 3.7: Distributions of (a) total integrated luminosity delivered by the LHC and recorded by ATLAS and (b) the peak instantaneous luminosity delivered by the LHC as a function of time in the 2010 data taking period.

for each lumi-block a set of indicators called data quality flags are used to identify beam and detector conditions. This information is exercised by physics analysers who use "Good Run Lists" (GRLs) to specify which quality flags they require to be passed for their analysis. For the purpose of this study, only runs passing certain quality requirements/flags were taken into account. In particular the following flags were required to be passed:

- ATLGL- requires that the run has been evaluated by the data quality group.
- LUMI- requires luminosity and forward detectors are operational and luminosity is correctly calculated.
- ATLSOL, ATLTOR- requires the currents in the solenoid and toroid magnets to be stable.
- L1CTP, L1CAL, L1MUB, L1MUE- requires the Level 1 Central Trigger Processor, calorimeter trigger and barrel/end cap muon triggers are performing with a good efficiency and there are no timing, synchronisation or data flow problems.
- cp_eg_electron_barrel, cp_eg_electron_endcap, cp_mu_mstaco, cp_mu_mmuidcb, cp_jet_jetb, cp_jet_jetea, cp_jet_jetec- are required to allow measurement of electrons, muons and jets. cp_met_metcalo, cp_met_metmuon- are required for $E_{\mathrm{T}}^{\mathrm{miss}}$ measurement (the latter was not required for periods A-C since this flag was not present).
- cp_tracking- requires inner detector in good working order.

- cp_btag_life- requires detector sub-systems required for b-tagging measurements are operable (not required for periods A-C since this flag was not present).

This was implemented by applying a GRL, based on that detailed in [40].

The 2010 data was analysed using the recommended physics containers, corresponding to datasets for each trigger stream. In order to maximise the amount of collected data different trigger requirements were made according to the running conditions in each run. The triggers used on the data are shown in Table 3.7 and taken from [40].

Electron Channel	
Period	Trigger
A-E3	L1_EM14
E4-I2	EF_e15_medium

Muon Channel	
Period	Trigger
A-E3	L1_MU10
E4-G1	EF_MU10_MG
G2-I1(< below run 167607)	EF_MU13_MG
I1(>= run 167607) - I2	EF_MU13_MG_tight

Table 3.7: Triggers used in the data for the different data running periods

Using this setup a total of $33.4\,\mathrm{pb}^{-1}$ of the 2010 data was analysed. A breakdown of how much luminosity was analysed per period is shown in Table 3.8.

Period	Dates	Runs	Recorded	Analysed
A	2010-Mar-30: 2010-Apr-19	152166-153200	$0.4\,\mathrm{nb}^{-1}$	$0.36\,\mathrm{nb}^{-1}$
B	2010-Apr-23: 2010-May-17	153565-155160	$9\,\mathrm{nb}^{-1}$	$7.8\,\mathrm{nb}^{-1}$
C	2010-May-17: 2010-Jun-05	155228-156682	$9.5\,\mathrm{nb}^{-1}$	$8.1\,\mathrm{nb}^{-1}$
D	2010-Jun-24: 2010-Jul-19	158045-159224	$0.32\,\mathrm{pb}^{-1}$	$0.24\,\mathrm{pb}^{-1}$
E	2010-Jul-29: 2010-Aug-18	160387-161948	$1.14\,\mathrm{pb}^{-1}$	$0.9\,\mathrm{pb}^{-1}$
F	2010-Aug-19: 2010-Aug-30	162347-162882	$2\,\mathrm{pb}^{-1}$	$1.65\,\mathrm{pb}^{-1}$
G	2010-Sep-22: 2010-Oct-07	165591-166383	$9.1\,\mathrm{pb}^{-1}$	$5.5\,\mathrm{pb}^{-1}$
H	2010-Oct-07: 2010-Oct-18	166466-166964	$9.3\,\mathrm{pb}^{-1}$	$6.8\,\mathrm{pb}^{-1}$
I	2010-Oct-24: 2010-Oct-29	167575-167844	$23\,\mathrm{pb}^{-1}$	$18.3\,\mathrm{pb}^{-1}$
Total	-	-	$\approx 46\,\mathrm{pb}^{-1}$	$\approx 33.4\,\mathrm{pb}^{-1}$

Table 3.8: Details of the data used in this analysis.

3.6 Conclusions

In this chapter the physical characteristics of the signal and backgrounds to its search were explored. The signal features include two hard, relatively isolated leptons, large missing energy and two typically forward high p_{T} jets. The signal is expected to suffer from a variety of backgrounds including $\mathrm{t\bar{t}}$ and single top, diboson production (WW,WZ and ZZ), W and Z production and QCD heavy quark and dijet production. In order to be able to maintain a good search sensitivity, given the small predicted cross section of Higgs production in the VBF mode, a very large background suppression will be required. In order to do this, the properties of the signal must be efficiently reconstructed within the ATLAS detector. This is explored in the next chapter.

Chapter 4

Physics Object Reconstruction

This chapter outlines the reconstruction and identification methods used in ATLAS by which raw signals in the sub-detectors are made to form meaningful physics objects (e.g. electrons, muons, jets and missing energy). Specific corrections and modifications to this standard procedure used in this study are defined and where appropriate motivated. Object identification efficiencies and mis-identification rates are calculated and a comparison between Monte Carlo simulation and 2010 data of important variables is made to ensure the objects are well understood.

4.1 Reconstruction of physics objects

The aim of the "reconstruction" of physics objects is to as accurately as possible reconstruct, on an event by event basis, the truth particles within the Monte Carlo simulation and the actual particles produced in real proton-proton collisions from the raw signals from the sub-detectors. Within ATLAS, this procedure is performed within the ATHENA framework (version 15.6.13 was used in this study). Because of limitations in the measurement accuracy of the detector sub-systems or approximations in the algorithms used to perform the reconstruction process, a 100 percent accurate reconstruction is not achievable. For example a particle may be reconstructed as the wrong type or not be detected at all. In order to quantify the level of this inconsistency the definitions of reconstruction efficiency and misidentification rate are commonly used. The reconstruction efficiency ϵ of a particle represents the fraction of true particles (in this study corresponding to the Monte Carlo truth particles) that are correctly reconstructed as that type of physics object. In this case the reconstruction efficiency can be expressed as shown in Eqn. 4.1,

$$\epsilon = \frac{N^{matched}}{N^{all}} \tag{4.1}$$

where N^{all} is the total number of Monte Carlo truth particles and $N^{matched}$ is the number of these that *match* a reconstructed object of the same type.

In order to quantify the rate at which a particle is mis-identified, the mis-identification rate (χ) shown in Eqn. 4.2 is defined as the number of reconstructed objects not matched to a truth particle of the same type ($N^{not-matched}$) divided by the total number of reconstructed objects ($N^{all-reconstructed}$).

$$\chi = \frac{N^{not-matched}}{N^{all-reconstructed}} \tag{4.2}$$

The signal within this analysis contains electrons, muons and jets and so it is particularly important that these objects are well reconstructed. To this end, the efficiency and mis-identification rates in the signal have been studied and compared to the performance of some of the main backgrounds in this analysis. The matching criteria used is a geometrical matching requiring the ΔR between the truth particle and reconstructed object to be less than 0.02 for electrons and muons and 0.1 for jets.

First the methods used to reconstruct these objects are discussed and any specific requirements for the objects used in the analysis are defined.

4.2 Electrons

The high rates expected from the vast QCD background at the LHC will make it difficult to correctly reconstruct and identify electrons over the broad p_{T} range they will be produced in by the physics channels of interest. Within the p_{T} range 20-50 GeV/c, the rate of production of isolated electrons compared to QCD jets will be below 10^{-5}. Although this effect is reduced at higher energy, high jet-rejection is required.

Electron Reconstruction Electrons are reconstructed using both calorimeter and inner detector information in ATLAS. There are two main algorithms for reconstruction within the inner detector acceptance ($|\eta| < 2.5$). The EGAMMA algorithm is designed to reconstruct high E_{T} isolated electrons. It is seeded by energy deposits in cells in the electromagnetic calorimeter and then searches for a matching track in the inner detector. The $softe$ algorithm is optimized to reconstruct "$soft$" (low p_{T}) electrons and is seeded by an inner detector track. It then searches for a matching EM cluster in the electromagnetic calorimeter. The electrons in the H $\rightarrow$ ZZ $\rightarrow \ell\ell\nu\nu$ signal are expected to be energetic so only electrons reconstructed with the EGAMMA algorithm are used and are discussed in the following.

Cluster reconstruction within the electromagnetic calorimeter is done using the *SlidingWindow* algorithm [62]. With the calorimeter divided into elements of size $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$, the total energy of all cells in all longitudinal layers is first calculated for each element. A window with a fixed size of N×N elements in $\eta - \phi$ space is then moved across the calorimeter in steps of $\Delta\eta$ and $\Delta\phi$. If the $\sum E_{\mathrm{T}}$ of elements within the window is a local maximum and > 2.5 GeV a cluster is formed. The size of the window used depends on the particle being reconstructed and its location in the calorimeter. For electrons in the barrel, a window size of 3×7 clusters is used because their bending in the magnetic field (compared to photons) causes them to radiate soft photons along ϕ resulting in a wider shower shape. In the end-caps, the same cluster size of 5×5 is used for both electrons and photons because the effect of the magnetic field is less important. For each reconstructed cluster an inner detector track is searched for within a window of $\Delta\eta \times \Delta\phi = 0.05 \times 0.1$. For electrons with a cluster energy divided by the track momentum less than 10, identification cuts are then applied.

Electron Identification After a candidate electron has been reconstructed, an electron identification procedure is applied to establish its reconstruction quality. Currently the default identification procedure applies a series of cuts related to shower shape, tracking and cluster-track matching variables (which are optimized according to E_{T} and η). Three standard electron definitions are used: *Loose*, *Medium* and *Tight*, each corresponding to an increasingly selective set of cuts (see Table 4.1), whereby each definition includes the cuts of the looser definitions.

Loose electron identification is based on calorimeter only information. Cuts are imposed on the shower shape in the second layer of the electromagnetic calorimeter and hadronic leakage (amount of energy reaching the hadronic calorimeter). This identification gives a high electron identification efficiency of approximately 90 % but a small background rejection rate of 560 [29], meaning 1 in 560 reconstructed jets, normalised with respect to the number of particle jets reconstructed, is mis-identified as an electron.

For *Medium* electrons additional cuts are applied on the energy deposited in the first (strips) layer of the electromagnetic calorimeter, to provide rejection of $\pi^0 \to \gamma\gamma$ and hadron showers. Further to this requirements are made on tracking variables, namely number of hits in the silicon (pixels and SCT) detectors and the transverse impact parameter. *Medium* identification cuts provide increased jet rejection, by a factor of 3-4 compared to *Loose* selection, but worsen identification efficiency by 10 %, down to 80 %.

Tight cuts use all information available for electron reconstruction. In addition to *Medium* cuts, they include a hit in the b-layer of the pixel detector in order to

Type	Description	Variable name
Loose cuts		
Detector acceptance	$\|\eta\| < 2.47$	
Hadronic leakage	Ratio of ET in the first layer of the hadronic calorimeter to ET of the EM cluster (used over the range $\|\eta\| < 0.8$ and $\|\eta\| > 1.37$)	Rhad1
	Ratio of ET in the hadronic calorimeter to ET of the EM cluster (used over the range $0.8 < \|\eta\| < 1.37$)	Rhad
Second layer of electromagnetic calorimeter	Ratio in of cell energies in 3×7 versus 7×7 cells	R_η
	Lateral width of the shower	$w_{\eta 2}$
Medium cuts		
First layer of EM calorimeter	Total shower width	wstot
	Ratio of the energy difference associated with the largest and second largest energy deposit over the sum of these energies	Eratio
Track quality	Number of hits in the pixel detector (≥ 1)	
	Number of hits in the pixels and SCT (≥ 7)	
	Transverse impact parameter (< 5 mm)	d0
Track matching	$\Delta\eta$ between the cluster and the track (< 0.01)	$\Delta\eta 1$
Tight cuts		
b-layer	Number of hits in the b-layer (≥ 1)	
Track matching	$\Delta\phi$ between the cluster and the track (< 0.02)	$\Delta\phi 2$
	Ratio of the cluster energy to the track momentum	E/p
	Tighter $\Delta\eta$ cut (< 0.005)	$\Delta\eta 1$
Track quality	Tighter transverse impact parameter cut (<1 mm)	d0
TRT	Total number of hits in the TRT	
	Ratio of the number of high-threshold hits to the total number of hits in the TRT	
Conversions	Electron candidates matching to reconstructed photon conversions are rejected	

Table 4.1: Definition of cuts used in identification of *Loose*, *Medium* and *Tight* electrons.

reject electrons from photon conversions and hits in the TRT. Further requirements are placed on the ratio of TRT high-threshold hits to low-threshold hits in order to reject the dominant backgrounds from charged hadrons, and strict matching between the cluster and the extrapolated track in $\eta \times \phi$ space.

The *Loose*, *Medium* and *Tight* electron identification criteria for the $m_H = 200\,\mathrm{GeV}/c^2$ signal sample are compared by determining the reconstruction efficiencies in each case. The electron efficiency as a function of (a) cluster E_T and (b) η is shown in Figure 4.1. Reconstructed and generator leptons have $|\eta| < 2.5$ and $E_T > 20(22)\,\mathrm{GeV}$ (reconstructed(generated)). The difference in E_T is to take into account resolution effects. The relationship between reconstruction efficiency between *Loose*, *Medium* and *Tight* selections remains approximately constant in p_T and η. For electrons from the VBF $H \to ZZ \to llll$ signal, the average reconstruction efficiency is 78 %, 76 % and 63 % for the *Loose*, *Medium* and *Tight* identification definitions respectively.

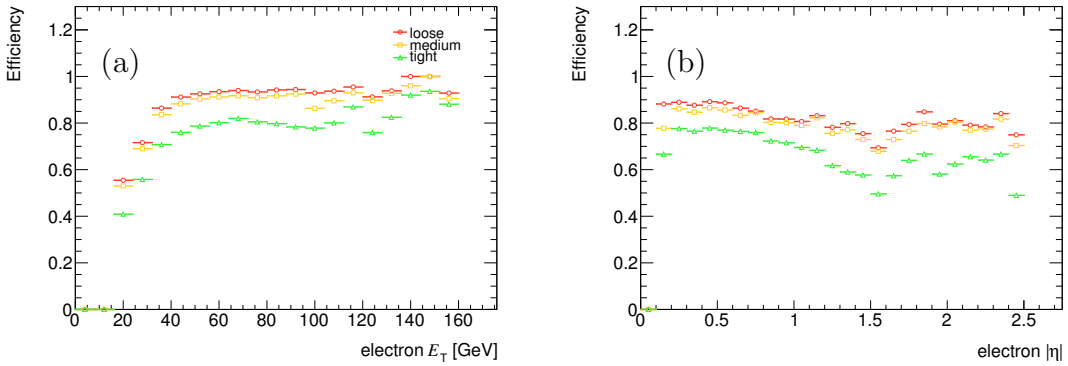


Figure 4.1: Comparison of *Loose*, *Medium* and *Tight* electron identification definitions in terms of electron reconstruction efficiency as a function of (a) calorimeter E_{T} and (b) η using VBF events in $\mathrm{m_H} = 200\,\mathrm{GeV}/c^2$ Higgs sample. The circular, square and triangular markers show the performance of the *Loose*, *Medium* and *Tight* criteria respectively.

4.2.1 Electrons in this analysis

Electrons used in this analysis are reconstructed with the standard EGAMMA algorithm and are required to have an electron track within the acceptance of the tracker and electromagnetic calorimeter and electron energy measured in the calorimeter $E_{\mathrm{T}} > 20\,\mathrm{GeV}$. Electrons are required to pass the *Robust-Medium* identification cuts. *Robust-Medium* electrons correspond to *Medium* electrons as defined in Section 4.2, with a few changes that were introduced by the $e\gamma$ performance group in order to maintain the robustness of electron identification by accounting for discrepancies between Monte Carlo and 2010 data. In particular the electron shower shapes were shown to be wider in data compared to Monte Carlo and as such cuts on R_η and $w_{\eta 2}$ are loosened. Further the hadronic leakage cut is modified due to a change in the modelling of the hadronic calorimeter noise (in early data it is modelled with a wider double gaussian). Any reference to electrons now refers to the *Robust-Medium* definition [63].

During the 2010 data taking period a number of cells in the electromagnetic calorimeter were lost because of problems in its readout electronics. In order to account for this effect, which is not taken into account in the Monte Carlo samples used, the (OTX) procedure to remove any electrons in the regions around the lost cells is implemented. In order to identify the lost cells the database corresponding to that at the end of the 2010 data taking period (from Run 167521) is used.

A number of additional corrections [64] are applied to the data and Monte Carlo which have been recommended by the e/γ performance group in order to improve the agreement between them. The energy scale of electrons in data is corrected by the expression $E_{Corrected} = E_{Original}/(1 + scale)$ where *scale* is equal to -0.0096

	Efficiency(%)		
Cut	VBF	$Z \to ee$	$t\bar{t}$
EGAMMA	90±1	95.5±0.1	72.1±0.2
robust Medium	79±1	86.6±0.1	50.1±0.2
OTX	72±1	78.9±0.1	45.7±0.2
$\|\eta\|$	72±1	78.9±0.1	45.7±0.2
E_T	71±1	77.6±0.1	44.9±0.2

Table 4.2: Efficiency of electron selection cuts. Efficiencies are shown for $m_H = 200\,GeV/c^2$ VBF $H \to ZZ \to llνν$ signal sample, $Z \to ee$ and $t\bar{t}$. Efficiencies are calculated by comparison of Monte Carlo truth and reconstructed electrons.

if $|\eta_{cluster}| < 1.4$ and 0.0189 for $1.4 < |\eta_{cluster}| < 2.5$, where $\eta_{cluster}$ is the electron cluster η. In order to maximise the reconstruction efficiency electrons within the crack region between $1.37 < |\eta| < 1.52$ are used. For the $m_H = 200\,GeV/c^2$ signal this was shown to provide an increase in VBF statistics by approximately 8 %. The energy of electrons in the crack is scaled by 5 % in the data and 3 % in the Monte Carlo. In addition, the identification efficiency of electrons measured in $Z \to ee$ and $W^{\pm} \to e\nu$ suggests that for *Robust-Medium* electrons Monte Carlo overestimates the efficiency. This is corrected for by weighting the Monte Carlo with η dependent scale factors [65].

Overall performance of electron identification with corrections The efficiency of the electron selection used in this analysis for the VBF $m_H = 200\,GeV/c^2$ signal sample and two different backgrounds which are expected to be a source of relatively isolated electrons ($Z \to ee$) and non-isolated electrons ($t\bar{t}$) are shown in Table 4.2. The efficiency is measured by comparing reconstructed electrons with respect to Monte Carlo truth level electrons with $E_T > 20\,GeV$ and $|\eta| < 2.5$, using a matching criteria of $\Delta R < 0.1$. As expected the source of isolated electrons $Z \to ee$ has the largest efficiency for the EGAMMA algorithm. However, it is promising that the signal shows a larger efficiency compared to the source of relatively non-isolated electrons, $t\bar{t}$, by nearly 20 %. Furthermore, requiring the *Robust Medium* identification, the efficiency for the signal and the $Z \to ee$ reduces by approximately 10 %, whereas for the $t\bar{t}$ this is closer to 20 %. Beyond this level, the cuts applied have a similar effect on each sample as expected.

A comparison of data and Monte Carlo of some basic electron variables including electron E_T and η (with/without crack electrons) is shown in Figure 4.2. As with all subsequent object distributions in this chapter, the plot shows the total background Monte Carlo distribution, represented by the total MC line together with the contribution of each background considered.The dataset analysed is added and shown by the circular markers. These distributions are made with the lepton requirements detailed in Section 5.2 of Chapter 5. Good agreement is found between data and

Monte Carlo for non-crack electrons while despite the corrections applied for crack electrons, they are underestimated at low E_{T} in the Monte Carlo.

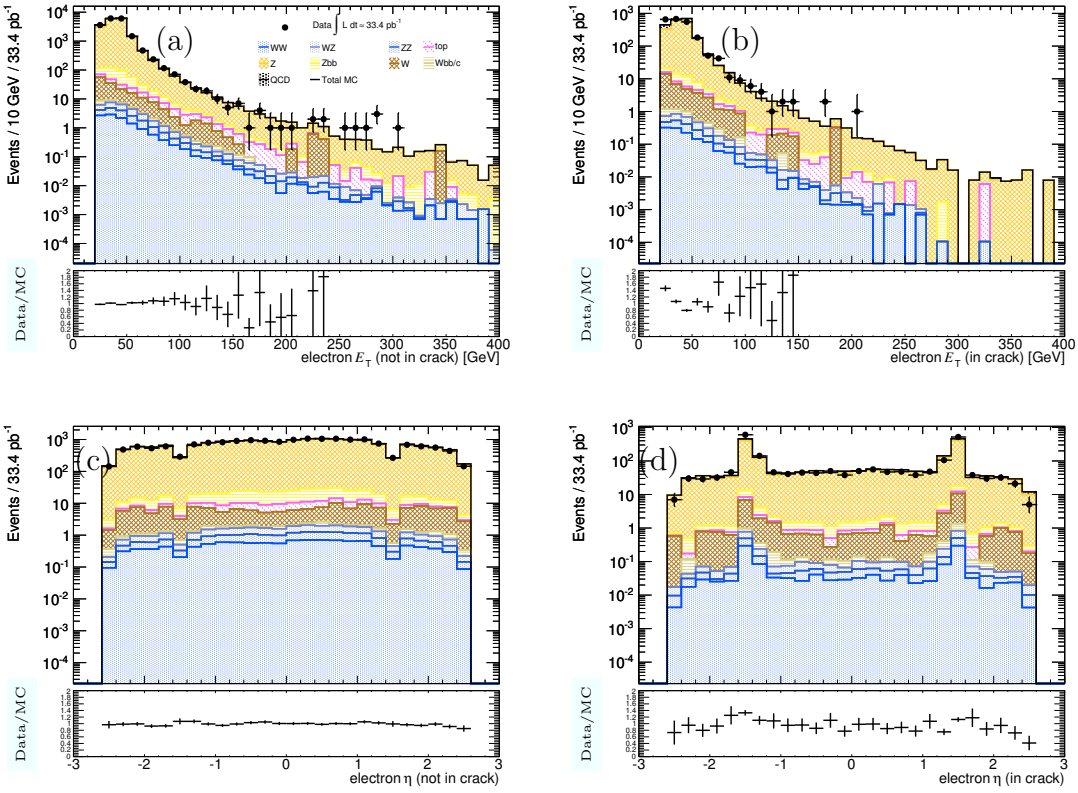


Figure 4.2: Comparison of data and Monte Carlo E_{T} and η distributions for electrons used in this analysis. The leftmost plots (not in crack) corresponds to events with none of the selected electrons in the crack. The rightmost (in crack) plots correspond to events with at least one selected electron lying in the crack region.

The electron identification efficiency and mis-identification rates as a function of electron E_{T} and $|\eta|$ are shown in Figure 4.3. Included is a comparison of the $\mathrm{m_H} = 200\,\mathrm{GeV}/c^2$ signal, $\mathrm{Z} \to ee$ and $\mathrm{t\bar{t}}$. Reconstructed electrons are required to pass the selection criteria outlined previously while Monte Carlo truth electrons must have $|\eta| < 2.5$ and $E_{\mathrm{T}} > 22(18)$ GeV for the efficiency(mis-identification) calculation respectively, in order to account for resolution effects. Electron efficiencies are observed to increase with E_{T}. With η the reconstruction efficiency is fairly constant although between $1.37 < |\eta| < 1.52$ it drops because of the electromagnetic barrel-end-cap transition. At higher η reconstruction efficiency worsens due to poorer tracking performance in the forward regions. Comparing the different samples included, it is seen that the source of non-isolated electrons $\mathrm{t\bar{t}}$ shows a lower reconstruction efficiency than the signal and the source of isolated electrons a higher efficiency. This trend is seen as a function of E_{T} and $|\eta|$.

The VBF component of the signal shows a larger mis-identification rate compared to the GF component. This is attributed to the VBF component being a source of

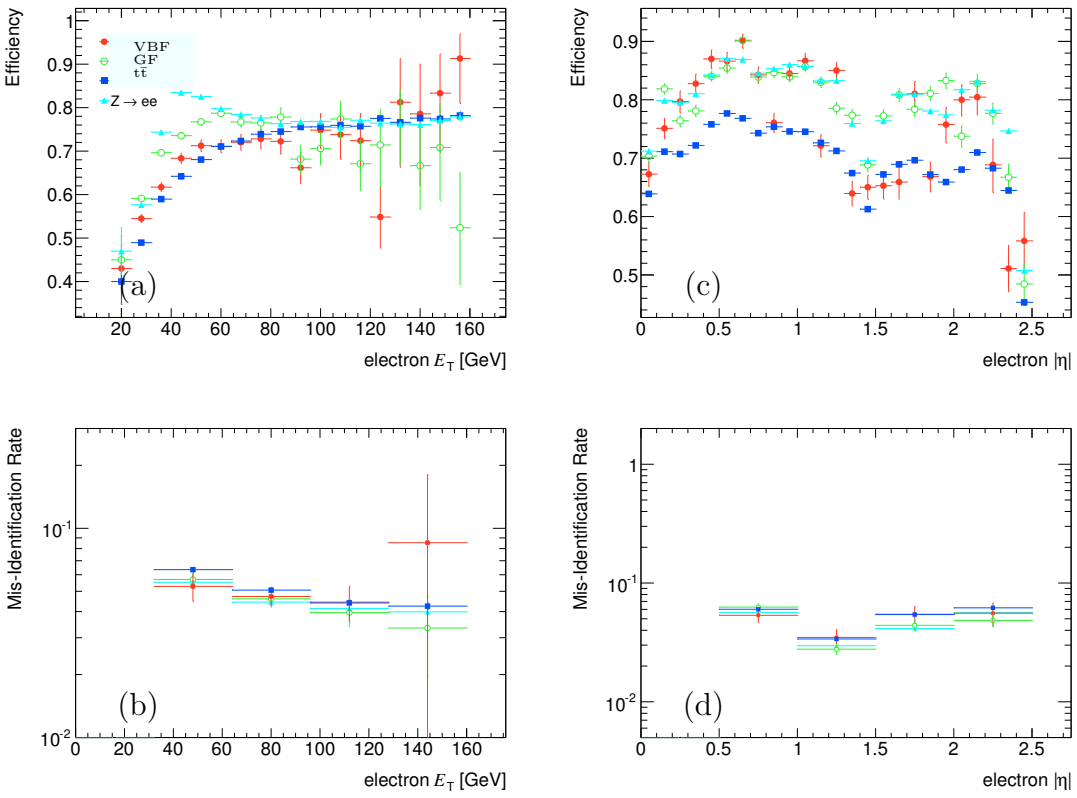


Figure 4.3: Electron reconstruction efficiency (upper plots) and mis-identification rate (lower plots) as a function of (a,b) E_{T} and (c,d) $|\eta|$ for VBF signal (filled circles), GF signal (open circles), $\mathrm{t\bar{t}}$ (filled squares) and $\mathrm{Z} \rightarrow ee$ (filled triangles). Signal components correspond to the $\mathrm{m_H} = 200\,\mathrm{GeV}/c^2$ sample.

less isolated electrons. For the same reason, the mis-identification rate for $t\bar{t}$ is larger than that for $Z \to ee$. As expected the mis-identification rates reduce with E_T due to the higher identification efficiency with increasing E_T. The trend in $|\eta|$ is approximately constant but is reduced in the crack region because of decreased efficiency.

4.3 Muons

Muons are reconstructed from the tracks they produce in the inner detector and muon spectrometer. Three different types of muon object can be reconstructed depending on the availability of information from these detector sub-systems. Standalone muons are reconstructed from a muon spectrometer track over its acceptance ($|\eta| < 2.7$). Combined Muons are reconstructed over the acceptance of the inner detector ($|\eta| < 2.5$), by matching a standalone muon to an inner detector track and combining the measurements. Segment tagged muons are reconstructed from an inner detector track matched to a short muon spectrometer track, typically within one innermost station (called a segment).

Within ATLAS there exist two algorithms to reconstruct each type of muon object. These algorithms are grouped into two families such that each family includes one algorithm for each type of reconstructed muon object. Each family is named by its corresponding combined reconstruction algorithm, and as such the families are STACO [66] and MUid [67].

Standalone Muons Standalone muon tracks in the muon spectrometer are reconstructed by joining track segments (straight lines of hits), built in each muon station. In the MUid-family of algorithms, the MOORE algorithm [68] makes the tracks while for the STACO-family this is done by the MuonBoy algorithm [69]. Subsequently the reconstructed muon spectrometer tracks are extrapolated to the beamline. This procedure takes into account both multiple scattering and energy loss in the calorimeter. Both MOORE and Muonboy do this by using a parametrisation of the energy loss and measuring the track quality with a χ^2 estimator. MOORE also takes into account energy measurements from the calorimeter to correct for the Landau tails of the energy loss distribution [70].

Combined Muons Combined muons [71] are formed by matching inner detector and muon spectrometer tracks corresponding to standalone muons. In both families a χ^2 (Eqn. 4.3) is used to measure the quality of matching and so decide which pairs of tracks are kept,

$$\chi^2_{match} = (\mathbf{M_{MS}} - \mathbf{M_{ID}})^T (\mathbf{C_{MS}} - \mathbf{C_{ID}})^{-1} (\mathbf{M_{MS}} - \mathbf{M_{ID}}) \tag{4.3}$$

where $\mathbf{M}_{MS/ID}$ are vectors of track parameters for the muon spectrometer standalone/ inner detector track respectively and $\mathbf{C}_{MS/ID}$ their corresponding covariance matrices. In the STACO algorithm, the resulting combined muon track is a statistical combination of the two tracks whereas the MUid algorithm performs a re-fit.

Segment tagged Muons Muons with insufficient momentum to traverse all muon stations are reconstructed by extrapolating inner detector tracks with large enough momentum to the inner/ middle muon stations (depending on η) and looking for matching reconstructed segments. Two methods are used to determine the quality of matching, depending on the family. Within the STACO family, track-segment matching quality is performed by the MuTag algorithm [66] using a χ^2 whereas in the MUid family, the MuGirl algorithm [72] uses an artificial neural network.

4.3.1 Muons in this analysis

The muon reconstruction efficiency using all combined and segment tagged STACO muons was compared to that obtained using just combined STACO muons. This was done using the $m_H = 200\,\mathrm{GeV}/c^2$ VBF signal as a function of (a) p_T and (b) η, as shown in Figure 4.4. Reconstructed and Monte Carlo truth muons with $|\eta| < 2.5$ and $p_T > 20(22)\,\mathrm{GeV}/c$ ((reconstructed)(Monte Carlo truth)) were used. By including segment tagged muons the global muon reconstruction efficiency is improved from 92 % to 95 % in this sample.

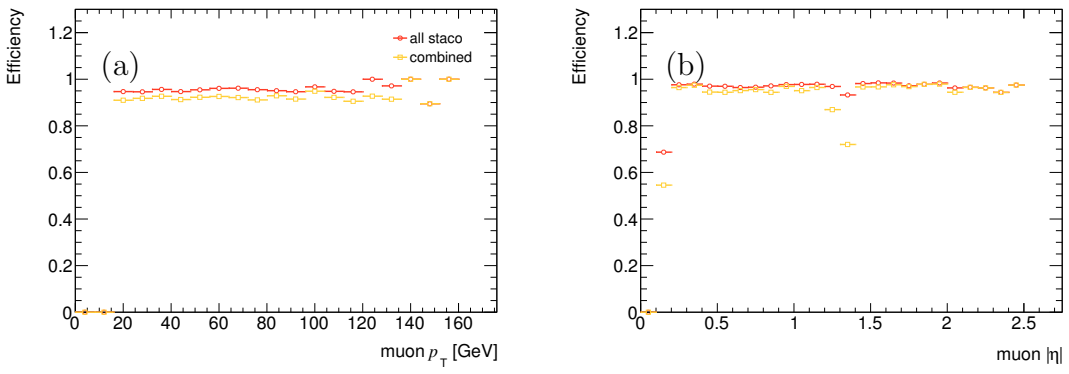


Figure 4.4: Muon reconstruction efficiency as a function of (a) p_T (b) η using $m_H = 200\,\mathrm{GeV}/c^2$ Higgs sample. The circular and square markers show the performance of using combined and segment tagged STACO muons and just combined STACO muons respectively.

In Figure 4.5 a comparison of combined, segment tagged and standalone muons is made by plotting the difference between reconstructed and truth matched muon p_{T} in $\mathrm{Z} \to \mu\mu$ events. It is clear standalone muons have a poor resolution. For this reason standalone muons are not considered in this analysis.

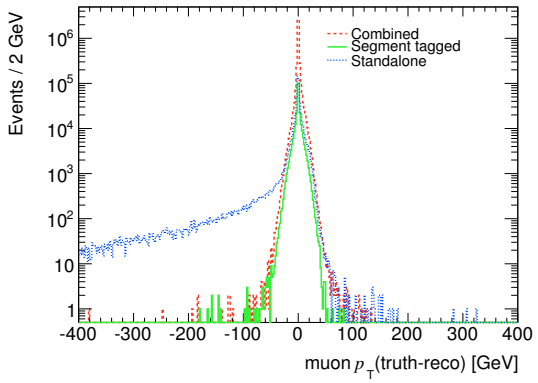


Figure 4.5: Comparison p_{T}(truth)-p_{T}(reco) of combined, segment- tagged and standalone muons. Monte Carlo sample used corresponds to $\mathrm{Z} \to \mu\mu$ events.

In this study, combined and segment tagged muons from the STACO family are used. Muons must have a $p_{\mathrm{T}} > 20$ GeV/c and $|\eta| < 2.5$. Muon tracks are required to be isolated by requiring the sum of the momenta of all tracks with $p_{\mathrm{T}} > 1$ GeV/c in a cone $\Delta R = 0.2$ around the muon to be less than 0.8 GeV/c. Additional requirements are made on the muon quality subject to the current recommendations from the muon combined performance group guidelines [73] which are largely to protect against muon solenoid - inner detector track mis-matching. As such identified muons are required to have traversed all the inner detectors sub-detectors, depositing ≥ 1 pixel hits, ≥ 6 SCT hits on the muon track. Within the acceptance of the TRT, hits are required as follows. Defining $\mathrm{n}_{TRThits}$ as the number of TRT hits on the muon track and $\mathrm{n}_{TRToutliers}$ the number of TRT outliers on the muon track, for $\eta < 1.9$ the sum of $\mathrm{n}_{TRThits}$ and $\mathrm{n}_{TRToutliers}$ (n) is required to be greater than 5 and $\mathrm{n}_{TRToutliers} < 0.9\,\mathrm{n}$. For $\eta > 1.9$, if $\mathrm{n} > 5$, then it is required that $\mathrm{n}_{TRToutliers} < 0.9\,\mathrm{n}$. Additionally for combined muons χ^2_{match} is required to be less than 150 and for tracks with muon spectrometer transverse momentum $p_{\mathrm{T}}(\mathrm{MS}) < 50$ GeV/c the difference between the extrapolated momentum in the muon spectrometer p(MS extrap) and the momentum in the inner detector p(ID) must be greater than $0.4p$(ID). In order to suppress cosmics the distance of closest approach relative to the primary vertex (the transverse impact parameter, d0) is required to be less than 1 mm and the absolute value relative to the z vertex at the beam-line (Z0) is required to be less than 1 cm.

A number of additional corrections are applied to muons in the Monte Carlo following the recommendations from the muon performance group. The performance of the transverse momentum scale and resolution for combined muons was measured in the data using the di-muon mass distribution in $Z \to \mu\mu$ decays [74]. It was found that the muon energy scale is reasonably well described and so no correction is applied. However, it was shown that the resolution in the data is poorer compared to that in the Monte Carlo. The difference is used to define smearing factors for the Monte Carlo for the inner detector and muon spectrometer momentum components. If both the inner detector and muon spectrometer components were measured, the overall transverse momentum of the combined muon is found by weighting the components by their relative resolution. For cases where no measurement was made in the muon spectrometer(inner detector) the muon p_T is taken to be the smeared inner detector(muon spectrometer) p_T. This procedure was done using the official code provided by the muon performance group. In [75] it is shown that for combined and segment tagged muons, the muon efficiency in the simulation and the data are well matched. Therefore no correction is applied.

Performance of muon reconstruction The efficiency of the muon selection used in this analysis for the $m_H = 200\,\mathrm{GeV}/c^2$ signal sample and two different backgrounds which are expected to be a source of relatively isolated ($Z \to \mu\mu$) and non-isolated ($t\bar{t}$) muons are shown in Table 4.3. The efficiency is measured by comparing reconstructed muons with respect to Monte Carlo truth level muons with $p_T > 20\,\mathrm{GeV}/c$ and $|\eta| < 2.5$, using a geometrical matching criteria of $dR < 0.1$. The efficiency achieved in the signal is close to that in $Z \to \mu\mu$ and similar to the value quoted in [75] of 97 %. The performance for $t\bar{t}$ is reduced by a further 10 % and is shown to be due to the track isolation cut, due to the presence of non-isolated muons in this sample.

Figure 4.6 compares data and the Monte Carlo prediction for some of the main muon variables. The distributions shown are made using the lepton requirements detailed in Section 5.2 of Chapter 5. A good level of agreement is found between the variables considered.

The muon identification efficiency and mis-identification rates for the muon selection adopted in this analysis as a function of muon p_T and $|\eta|$ are shown in Figure 4.7. Included is a comparison of the $m_H = 200\,\mathrm{GeV}/c^2$ signal, $Z \to \mu\mu$ and $t\bar{t}$ samples. Reconstructed muons must satisfy the muon selection detailed previously and Monte Carlo truth muons $|\eta| < 2.5$ and $p_T > 22(18)\,\mathrm{GeV}/c$ for the efficiency(mis-identification) calculation. Muon efficiencies are observed to increase with p_T until $60\,\mathrm{GeV}/c$ at which point there is a plateau region where they remain approximately constant. With η the reconstruction efficiency rises quickly from low η where the

	Efficiency(%)		
Cut	VBF	$Z \to \mu\mu$	$t\bar{t}$
Combined/segment tagged	99±2	99.0±0.1	99.1±0.3
Pixel hits	99±2	99.0±0.1	98.9±0.3
SCT hits	99±2	98.5±0.1	98.5±0.3
TRT hits	99±2	98.3±0.1	98.1±0.3
χ^2_{match}	98±2	98.0±0.1	97.8±0.3
p(ID)	98±1	97.6±0.1	97.3±0.3
d0	98±1	97.6±0.1	97.0±0.3
Z0	98±1	97.6±0.1	97.0±0.3
Track Isolation	95±1	95.9±0.1	80.9±0.3
p_T	94±1	94.9±0.1	80.3±0.3
$\|\eta\|$	94±1	94.9±0.1	80.3±0.3

Table 4.3: Efficiency of Muon selection cuts. Efficiencies are shown for $m_H = 200\,\mathrm{GeV}/c^2$ VBF $H \to ZZ \to llνν$ signal sample, $Z \to \mu\mu$ and $t\bar{t}$. Efficiencies are calculated by comparison of Monte Carlo truth and reconstructed muons.

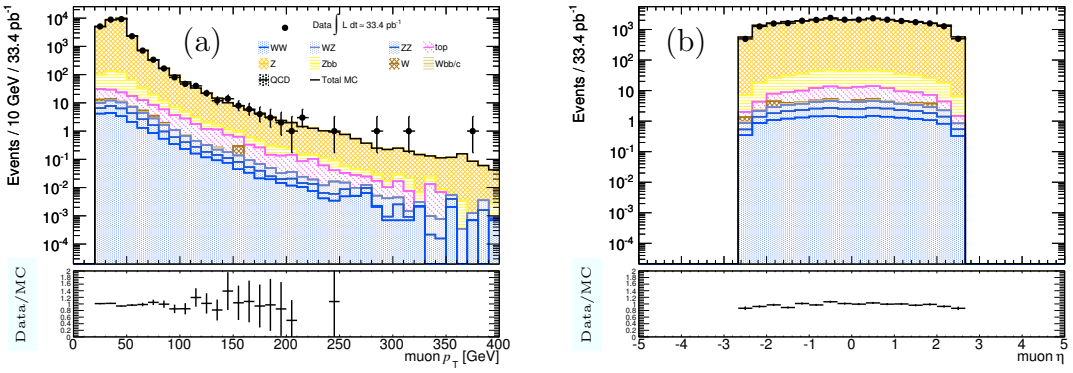


Figure 4.6: Distributions of muon (a) p_T and (b) $|\eta|$ for muons used in this analysis as detailed in Section 4.3.1.

detector acceptance is poor due to the detector support structure. Efficiency is also degraded in the transition region between $1.1 < |\eta| < 1.7$ where there are fewer muon stations. Muon mis-identification rates are highest for the $t\bar{t}$ due to the presence of non-isolated muons making reconstruction more difficult. At the other extreme, $Z \to \mu\mu$ exhibits the lowest mis-identification rate. Although higher than in the GF component of the signal, the mis-identification rate of the VBF signal is lower than that in $t\bar{t}$. These findings indicate that adopting the selection described, muons can be identified efficiently in the signal and with fewer mistakes compared to a major expected background, $t\bar{t}$.

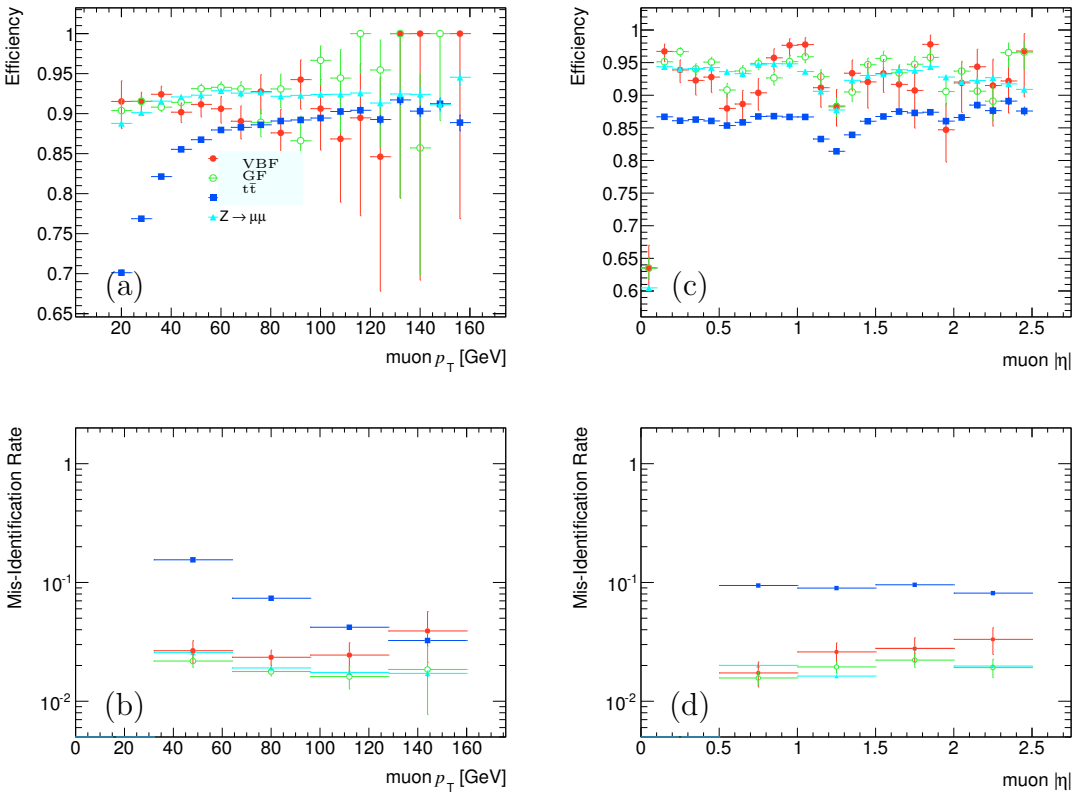


Figure 4.7: Muon reconstruction efficiency (upper plots) and mis-identification rate (lower plots) as a function of (a,b) p_T and (c,d) $|\eta|$ for VBF signal (filled circles), GF signal (open circles),$t\bar{t}$ (filled squares) and $Z \to \mu\mu$ (filled triangles) after the pre-selection of muons described in Section 4.3.1. Signal components correspond to the $m_H = 200\,\mathrm{GeV}/c^2$ sample.

4.4 Jets

Jets are collimated hadrons produced by energetic partons. They deposit energy in the electromagnetic and hadronic calorimeters and if charged, tracks in the inner detector. Jets are reconstructed using calorimeter information, over its full acceptance ($|\eta| < 4.9$). They are reconstructed using jet-finding algorithms, of which

there are many examples, but each of which reconstructs jets from input objects by combining their four momenta. The input to jet-finding algorithms does not have to be derived from the calorimeter (giving calorimeter jets), tracks or particles from an event generator are possible inputs. In order to reconstruct calorimeter jets, the calorimeter cells are combined into larger objects called calorimeter towers or topological cell clusters, which form the input to the jet-finding algorithms.

Input to jet finding Calorimeter towers are formed by projecting calorimeter cells onto a fixed grid in pseudo-rapidity and azimuth, over the whole calorimeter acceptance. The size of the towers is constant over all the calorimeter with dimensions $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$. Projective calorimeter cells completely contained within a tower contribute all of their energy to the tower whereas projective cells that extend beyond a tower boundary, share their energy across several towers in proportion to the level of overlap between the cell area in $\eta \times \phi$ and the tower.

Topological clusters represent the three-dimensional energy deposit derived from the reconstruction of showers developing in the calorimeter. Their formation is seeded by high energy deposits in a (seed) cell with signal-to-noise ratio $(E_{cell}/\sigma_{cellnoise})$ > 4. Neighbouring cells contribute if they have $(E_{cell}/\sigma_{cellnoise}) > 2$. Subsequently the cells with $(E_{cell}/\sigma_{cellnoise}) > 0$ are added to the cluster. This procedure corresponds to the 4/2/0 noise suppression procedure and stops when cells with low signal to noise ratio are found around the seed cell. Subsequently a splitting algorithm is run on each cluster to separate those with more than one local maximum.

Jet algorithms A number of jet finding algorithms exist to build the jets used in an analysis from calorimeters towers or topological clusters. The popular anti-k_T jet finding algorithm [76, 77, 78] calculates for each input object the smallest distance between objects (d_{ij}) and the distance between the object and the beam (d_{iB}), which are then stored in a list. The smallest valued variable is identified and if it corresponds to the variable d_{ij}, objects i and j are combined (i.e. 4-momenta are added) whereas if it is a d_{iB}, the object i is considered as a jet and is removed from the list. The variables are recalculated for each object and the procedure repeated until no objects remain. The distance parameter R is used to set the relative distance at which jets can be resolved from each other as compared to the beam.

Jet Calibration The calorimeters used in ATLAS are non-compensating, meaning that two particles which have the same energy and interact with the calorimeter, but one of which interacts electromagnetically and the other hadronically, will give rise to different cell signal densities in the calorimeter. Jets must be calibrated to account for this effect and in addition to account for detector effects, including noise,

losses in dead material and cracks and particle deflection in the magnetic field. Two methods exist in ATLAS to perform these calibrations. Global calibration involves calibrating jets once they have been built out of calorimeter objects. In contrast local-hadronic calibration calibrates calorimeter clusters.

Global hadronic calibration based on cell signal weighting was developed by the CDHS experiment and further refined by other experiments. Its purpose is to identify and appropriately weight calorimeter cells derived from hadronic interactions. Typically hadronic interactions produce low cell signal density compared to those from electromagnetic interactions, providing a means by which the associated cells can be identified and weighted of the order of e/π signal ratio to calibrate them. The cell weighting was derived from Monte Carlo simulations of di-jets and depends on the cell location and the cell signal density. In local hadronic calibration, the calibration is done before the jet building process and as such the resulting jets are already calibrated to the local hadronic energy scale. However, corrections are still needed to account for detector effects and the fact that the calibrations are derived from single pion response.

4.4.1 Jets used in this analysis

A comparison of using calorimeter towers and topological clusters as input to jets built with the anti-k_T algorithm (with global calibration) was made using VBF events in the $m_H = 200\,\mathrm{GeV}/c^2$ signal sample. This is shown in Figure 4.8. The uppermost figures compare the reconstruction efficiency obtained with the two inputs to jet finding as a function of p_T and $|\eta|$. Reconstructed(Monte Carlo truth) jets(quarks) were required to have $p_T > 20(22)\,\mathrm{GeV}/c$ and $|\eta| < 4.9$ and considered matched if $\Delta R < 0.3$. Both inputs give very similar performance. This is verified in the lower plots of Figure 4.8 which show the variable p_T(reco)-p_T(truth)/p_T(truth) (where p_T(reco/truth) refers to a reconstructed jet/ Monte Carlo truth quark p_T) as a function of the Monte Carlo truth quark p_T and $|\eta|$.

The jets used in this analysis are reconstructed using the anti-k_T algorithm with a distance parameter of 0.4, calibrated using global calibration. Topological clusters are used for the jet algorithm input. The energy scale of jets is corrected from the electromagnetic scale to the hadronic scale using the recommended p_T and η dependent Jet Energy Scale. This was derived from Monte Carlo but verified with data, and as such is expected to give a more accurate calibration at the time of writing. Jets must have a $p_T > 20\,\mathrm{GeV}/c$ and $|\eta| < 4.5$. In order to suppress the contribution of a jet in other than the p-p collision from which it originated, at least 75 % of any tracks originating from the jet are required to be associated with the primary vertex of the given p-p collision. This is achieved by requiring that

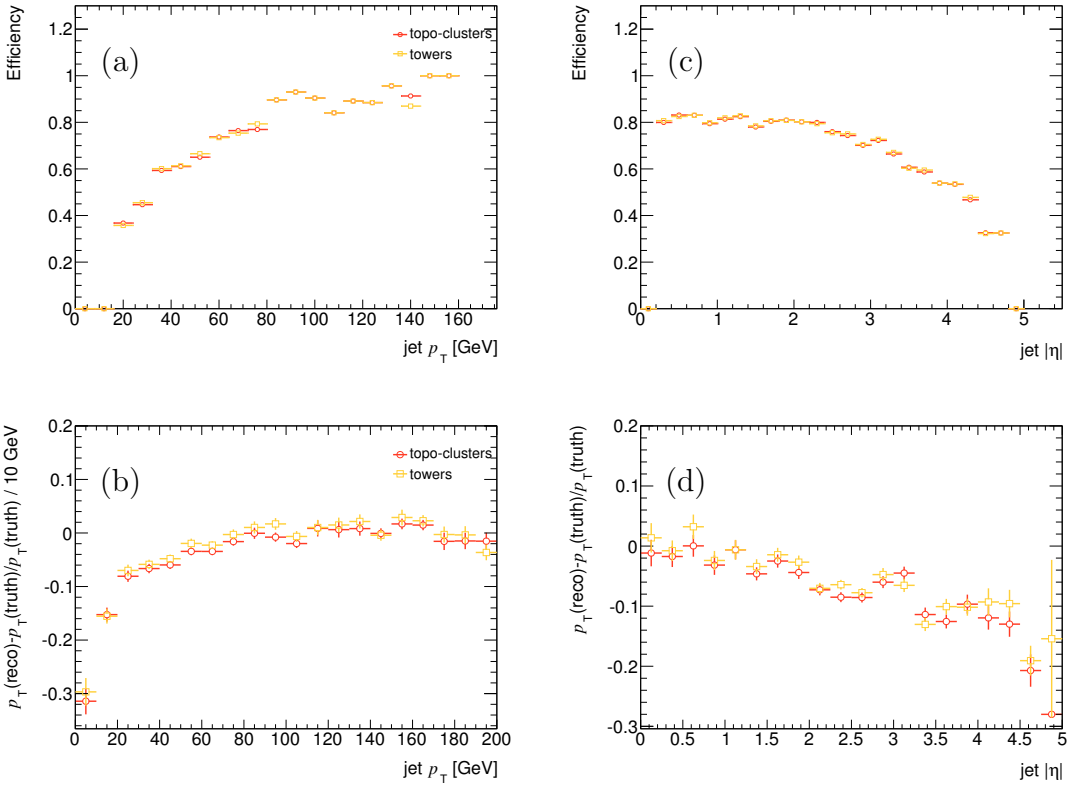


Figure 4.8: Jet reconstruction efficiency (upper plots) and profile of (p_{T}(reco)-p_{T}(truth)/p_{T}(truth) (lower plots) and as a function of (a,b) p_{T} and (c,d) $|\eta|$ in $\mathrm{H} \rightarrow ll\nu\nu$ events ($\mathrm{m_H} = 200\,\mathrm{GeV}/c^2$) comparing tower and topological clusters input to anti-k_T jets.

	Efficiency(%)	
Cut	VBF	$t\bar{t}$
p_T	96±1	95.5±0.1
$\|\eta\|$	96±1	95.5±0.1
JVF	96±1	95.4±0.1

Table 4.4: Efficiency of Jet selection cuts. Efficiencies are shown for $m_H = 200\,\mathrm{GeV}/c^2$ VBF $H \to ZZ \to llνν$ signal sample and $t\bar{t}$.

the jets have a Jet Vertex fraction |JVF| > 0.75. In this way the effect of in-time pile-up causing multiple p-p collisions to occur within the same bunch crossing can be reduced.

Within ATLAS there are a number of b-tagging algorithms to differentiate the decay of a b-hadron from that containing only light quarks. In this analysis this is particularly important in order to suppress backgrounds such as those with a top quark decay. The b-tagging algorithms use the fact that hadrons with a b quark have a much larger lifetime giving a pronounced decay length $c\tau \approx 450\,\mu\mathrm{m}$ compared to light quark hadrons. They are identified either by using reconstructed secondary vertices from the tracks within a jet or by combining the distance of closest approach to the primary vertex of all tracks in the jet. Within this analysis, decays of b hadrons are identified using the secondary vertex based algorithm SV0 [79]. A jet is called a b-jet if its lifetime-signed decay length significance (b-tag weight) is greater than 5.72 (this follows the recent $t\bar{t}$ analysis [49]).

A series of cuts are applied to ensure that the jets used are free from electromagnetic coherent noise bursts, calorimeter spikes and cosmics/ beam background. They follow the recommendation of the JetE_T^{miss} group [80] and as such any jet in the data which fails one of the *loose* criteria is rejected and not considered further in the analysis. Any event with a *bad* jet with $p_T > 20\,\mathrm{GeV}/c$ is then rejected. This selection is only implemented on data as the distributions relating to the selections are not well modelled in simulation.

Performance of jet reconstruction In Table 4.4, the efficiency of the jet selection used in this analysis for the $m_H = 200\,\mathrm{GeV}/c^2$ signal sample and $t\bar{t}$ samples are shown. The efficiency is measured by comparing reconstructed jets with respect to Monte Carlo truth quarks with $p_T > 20\,\mathrm{GeV}/c$ and $|\eta| < 4.5$, using a geometrical matching criteria of $\Delta R < 0.3$. It can be seen that the efficiency achieved in the signal is close to that in $t\bar{t}$.

A comparison of data and Monte Carlo for some important jet variables is shown in Figure 4.9 for jets passing the jet selection outlined. Agreement between data and Monte Carlo is reasonable apart from in the first bin in the p_T distribution where

the Monte Carlo under-estimates the data. It is possible this discrepancy is due to pile-up affecting low p_{T} events.

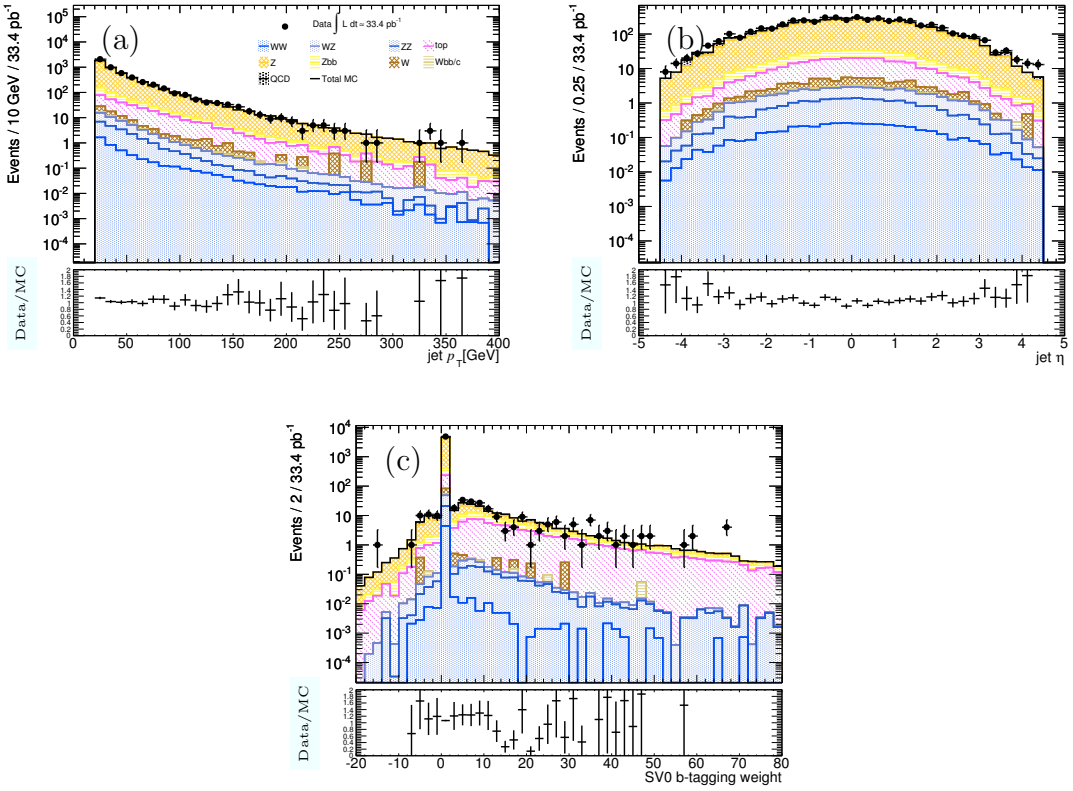


Figure 4.9: Distributions of jet (a) p_{T} (b) $|\eta|$ and (c) SV0 b-tag weight.

The reconstruction efficiency of jets used in this analysis as a function of p_{T} and $|\eta|$ is shown in Figure 4.10 for the $\mathrm{m_H} = 200\,\mathrm{GeV}/c^2$ signal, $\mathrm{Z} \to \mu\mu$+jets and $\mathrm{t\bar{t}}$ samples. The performance of jet reconstruction in the VBF signal appears to be slightly worse than in $\mathrm{t\bar{t}}$ at low jet p_{T} and large $|\eta|$. In comparison, the performance in $\mathrm{Z} \to \mu\mu$+jets is lower compared to that in the signal across the entire $|\eta|$ range investigated. Overall, for the signal, a jet reconstruction efficiency of over 60 % is maintained for jet $p_{\mathrm{T}} > 30\,\mathrm{GeV}/c$ and $|\eta| < 3.5$.

4.5 Missing energy ($E_{\mathrm{T}}^{\mathrm{miss}}$)

Typically weakly interacting particles such as neutrinos will traverse the entire detector volume without leaving any measurable signal. As such they have to be measured indirectly through the imbalance of observed transverse momentum they cause which is commonly expressed through the quantity $E_{\mathrm{T}}^{\mathrm{miss}}$.

Reconstruction of $E_{\mathrm{T}}^{\mathrm{miss}}$ in ATLAS [81, 82] is in essence done by summation of all energy deposits in the calorimeter cells and muon tracks. However it is complicated as there are many processes apart from the hard scattering process such as

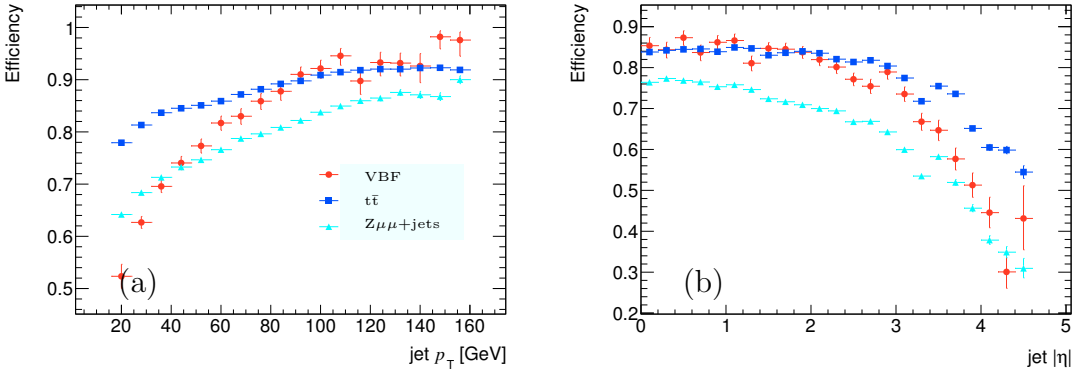


Figure 4.10: Jet reconstruction efficiency as a function of (a) p_{T} and (b) $|\eta|$ for VBF component of signal (filled circles), $\mathrm{t\bar{t}}$ (filled squares) and $\mathrm{Z} \rightarrow \mu\mu$+jets (filled triangles) after the preselection of jets described in Section 4.4.1. The signal sample corresponds to $\mathrm{m_H} = 200\,\mathrm{GeV}/c^2$.

the underlying event, multiple interactions, pile-up and coherent electronics noise, which give rise to such energy deposits and muon tracks. To achieve an accurate measurement of E_{T} and therefore $E_{\mathrm{T}}^{\mathrm{miss}}$, each component must be correctly calibrated and corrections for energy loss in non-active material have to be accounted for. It is further made difficult due to fake $E_{\mathrm{T}}^{\mathrm{miss}}$ coming from noisy/dead calorimeter cells, badly reconstructed/fake muons and acceptance effects due to lack of detector coverage.

Within ATLAS there are two methods to reconstruct $E_{\mathrm{T}}^{\mathrm{miss}}$, each with its own algorithm. The cell based method is used in this analysis and is described further in this section. The other method is object based and utilises reconstructed physics objects and any other low p_{T} objects not reconstructed to calculate the $E_{\mathrm{T}}^{\mathrm{miss}}$ through the imbalance of their combined total energy. It takes into account losses in non-active regions of the detector.

Cell based $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction

The total missing transverse energy in the cell based method, ${E_{\mathrm{T}}^{\mathrm{miss}}}^{Final}$ (Eqn. 4.4) is made up of contributions from the calorimeter (${E_{\mathrm{T}}^{\mathrm{miss}}}^{CaloCalib}$), the cryostat (${E_{\mathrm{T}}^{\mathrm{miss}}}^{Cryo}$) and the muons (${E_{\mathrm{T}}^{\mathrm{miss}}}^{Muon}$). In the following each is discussed.

$$ {E_{\mathrm{T}}^{\mathrm{miss}}}^{Final} = {E_{\mathrm{T}}^{\mathrm{miss}}}^{CaloCalib} + {E_{\mathrm{T}}^{\mathrm{miss}}}^{Cryo} + {E_{\mathrm{T}}^{\mathrm{miss}}}^{Muon} \tag{4.4} $$

Calorimeter contribution: ${E_{\mathrm{T}}^{\mathrm{miss}}}^{Calo}$ The 200k readout channels of the ATLAS calorimeter give rise to electronics noise causing approximately a 13 GeV increase in width to the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution. The first step in the calculation of the

calorimeter term ($E_{\mathrm{T}}^{\mathrm{miss}Calo}$) involves the application of one of two methods to suppress this noise.

Standard Noise Suppression considers calorimeter cells with energies greater than a threshold. In most cases a symmetric threshold with $|E_{cell}| > n \times \sigma_{noise}$, where σ_{noise} is the width of the noise distribution and $n = 2$ is used.

Noise Suppression with TopoClusters considers calorimeter cells called TopoCells that lie within 3-dimensional topological clusters called TopoClusters. These TopoClusters are constructed using the 4/2/0 noise suppression procedure outlined in Section 4.4. Because of its linearity and resolution, this method is the default noise suppression method to be applied in both the cell and object based $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction schemes.

After application of noise suppression, the $E_{\mathrm{T}}^{\mathrm{miss}Calo}$ contribution is calculated using the x and y components of selected TopoCells, combined as in Eqn. 4.5,

$$\begin{aligned} E_x^{\mathrm{miss}Calo} &= -\sum_{i=1}^{N_{TopoCell}} E_i \sin\theta_i \cos\phi_i \\ E_y^{\mathrm{miss}Calo} &= -\sum_{i=1}^{N_{TopoCell}} E_i \sin\theta_i \sin\phi_i \\ E_{\mathrm{T}}^{\mathrm{miss}Calo} &= \sqrt{\left(E_x^{\mathrm{miss}Calo}\right)^2 + \left(E_y^{\mathrm{miss}Calo}\right)^2} \end{aligned} \tag{4.5}$$

where $N_{TopoCell}$ is the number of cells satisfying the noise suppression, E_i is the ith TopoCell energy and θ_i and ϕ_i the corresponding azimuthal and polar angle.

This process results in a large shift in $E_{\mathrm{T}}^{\mathrm{miss}}$ by approximately 30 % compared to true $E_{\mathrm{T}}^{\mathrm{miss}}$. A further calibration scheme is applied to remove this. In the first step energy deposits are classified as either electromagnetic, hadronic in origin or resulting from high p_{T} particles. Typically electromagnetic showers have a higher energy density compared to hadronic showers. In addition electromagnetic showers are normally wider than hadronic showers, depositing their energy in a smaller distance. Subsequently methods to calibrate hadronic showers, like H1 or local-hadronic calibration, as detailed in Section 4.4 are used to calibrate each calorimeter cell. The corresponding calibrated $E_{\mathrm{T}}^{\mathrm{miss}Calo}$ is called $E_{\mathrm{T}}^{\mathrm{miss}CaloCalib}$.

Muon contribution: $E_{\mathrm{T}}^{\mathrm{miss}Muon}$ is calculated from the momenta of reconstructed muons in $|\eta| < 2.7$ as in Eqn. 4.6.

$$E_{\mathrm{T}}^{\mathrm{miss}Muon} = -\sum_{\mathrm{reconstructed muons}} E_{\mathrm{T}} \tag{4.6}$$

Inside the acceptance of the inner detector $|\eta| < 2.5$, only muons with a matched inner detector track are taken into account. This reduces fake muon contributions occurring in events with energetic jets that cause large hit multiplicities in the muon spectrometer. The momentum of muons as measured in the muon spectrometer is used. No p_{T} threshold is required. Loss of muons due to gaps in the detector coverage are not currently recovered. The $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution is largely unaffected by the muon term. However, badly measured or fake muons can be a large source of fake $E_{\mathrm{T}}^{\mathrm{miss}}$.

Cryostat contribution: $E_{\mathrm{T}}^{\mathrm{miss}\,Cryo}$ The purpose of the cryostat term ($E_{\mathrm{T}}^{\mathrm{miss}\,Cryo}$) is to allow for the recovery of energy loss in the cryostat (barrel and end-caps) through hadronic showers. This is achieved by using the correlation between the energy in the outermost compartment of the electromagnetic calorimeter and the energy in the first layer of the hadronic calorimeter. It is defined below in Eqn. 4.7

$$
\begin{aligned}
E_{\mathrm{T}}^{\mathrm{miss}\,Cryo} &= -\sum_{recJets} Ejet_T^{Cryo} \\
Ejet_T^{Cryo} &= \omega^{Cryo}\sqrt{E_{EM3} \times E_{HAD}}
\end{aligned}
\tag{4.7}
$$

where ω^{Cryo} is a calibration weight (found using H1-like calibration) and E_{EM3} and E_{HAD} are the jet energies in the third compartment of the electromagnetic calorimeter and the first layer of the hadronic calorimeter respectively. This correction gives rise to a $\approx 5\,\%$ per jet shift for jets with p_{T} >500 GeV/c.

Definition of Missing Energy used in this analysis

The definition of $E_{\mathrm{T}}^{\mathrm{miss}}$ used in this analysis follows what at the time of writing was currently recommended. It is called $\mathrm{MET}_{LocHadTopo}$ and uses the cell-based approach detailed previously using TopoClusters within $|\eta| < 4.5$. The calorimeter cell energy is calibrated using weights from local hadron calibration of TopoClusters. As recommended, in events where there are muons passing the muon selection criteria outlined in Section 4.3.1, the muon terms ($E_{\mathrm{T}}^{\mathrm{miss}\,Muon}$) in the $E_{\mathrm{T}}^{\mathrm{miss}}$ definition are removed and replaced explicitly with the muons found in the event, in order to avoid double counting. The motivation for this is shown in Figure 4.11, where the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution in $\mathrm{Z} \to \mu\mu$ events is presented when different types of muons are selected (rows) and different types of muon are used in order to correct the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution (columns). The default $E_{\mathrm{T}}^{\mathrm{miss}}$ variable is plotted in the left-most column. Combined and segment tagged muons are required to have a $p_{\mathrm{T}} > 20\,\mathrm{GeV}/c$ and $|\eta| < 2.5$ while standalone muons are required to have $p_{\mathrm{T}} > 20\,\mathrm{GeV}/c$ and $|\eta| <$

2.7. When combined and segment tagged muons are used in the selection (like in this study) if no replacement of the muon terms is made in the $E_{\mathrm{T}}^{\mathrm{miss}}$ expression, then a bump around 50 GeV appears in the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution. This problem is only resolved when the muon terms are replaced explicitly with the combined and segment tagged muons in the event.

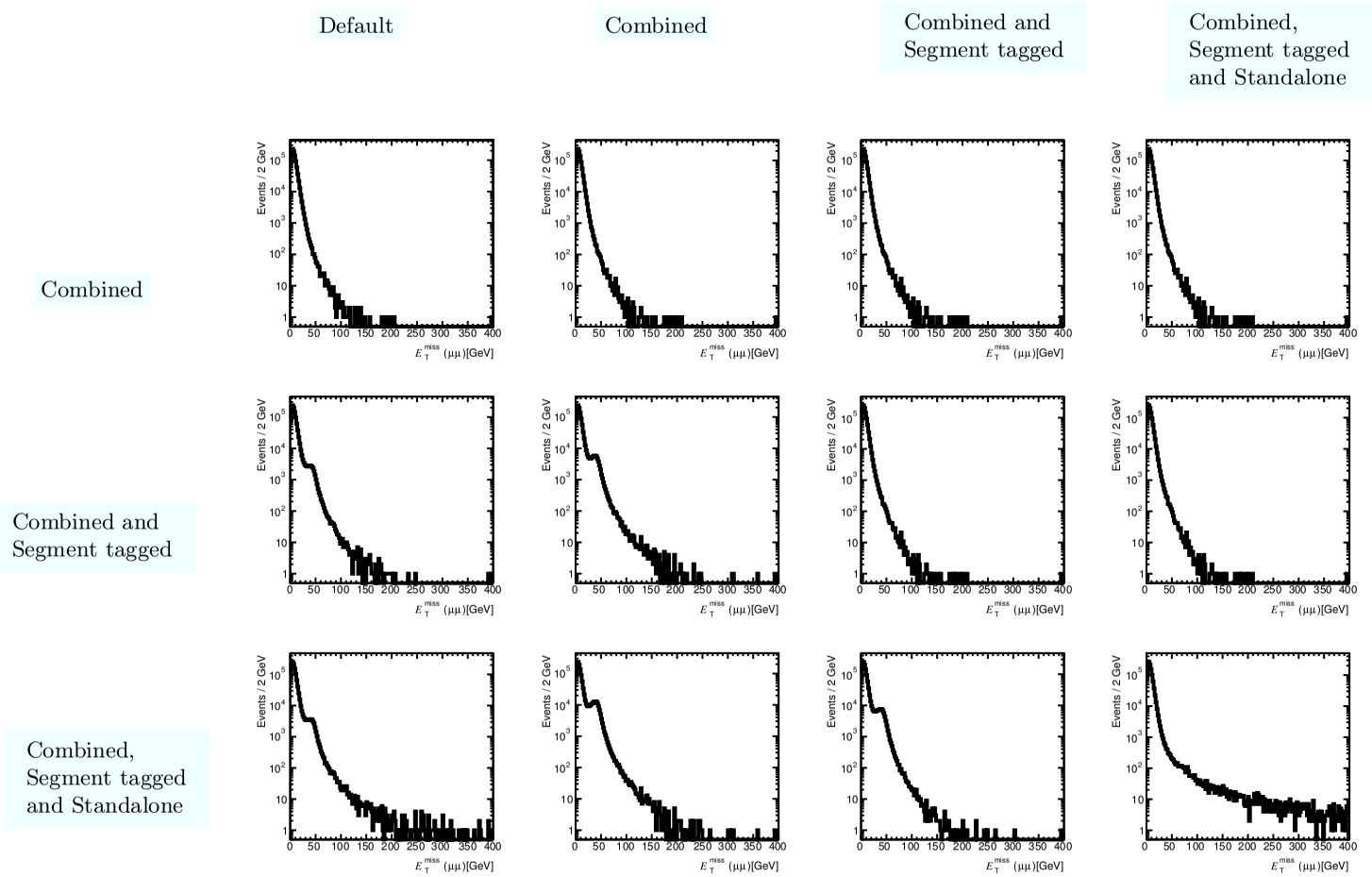


Figure 4.11: $E_{\mathrm{T}}^{\mathrm{miss}}$ in $\mathrm{Z} \to \mu\mu$ events. The rows represent the different types of muons considered. The leftmost column shows the default $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution. Moving from left to right in the columns, shows the effect of replacing the $E_{\mathrm{T}}^{\mathrm{miss}\,Muon}$ term with different types of muons.

This behaviour is further verified by Figure 4.12. Here, the difference between the Monte Carlo truth and reconstructed E_x^{miss},E_y^{miss} and $E_{\text{T}}^{\text{miss}}$ distributions, before and after the replacement of the muon terms for the $\text{Z} \to \mu\mu$ sample is shown.

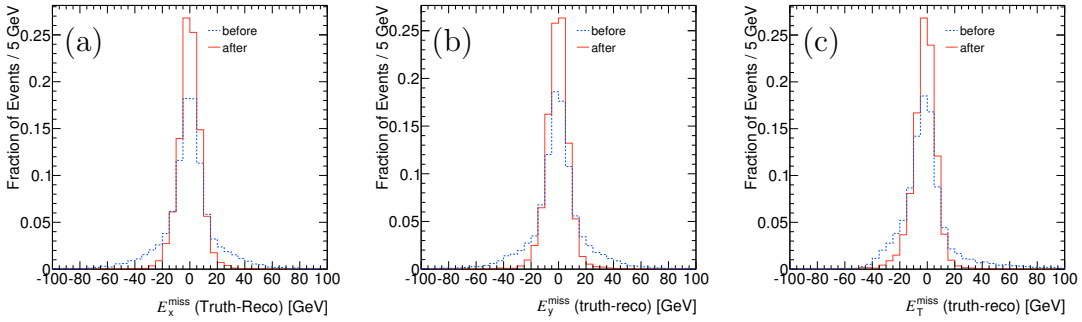


Figure 4.12: Comparison of reconstructed missing energy components (a) E_x^{miss}, (b) E_y^{miss} and (c) $E_{\text{T}}^{\text{miss}}$ with Monte Carlo truth before and after replacing the $E_{\text{T}}^{\text{miss}\,Muon}$ term in the $E_{\text{T}}^{\text{miss}}$ expression with combined and segment tagged muons found passing the muon selection.

4.6 Overlap removal

Within ATLAS an object satisfying the requirements to be reconstructed by the algorithm of different objects, will be recorded as each type of object. For example, reconstruction of jets is based on information from the calorimeter only. As a consequence of this, energy depositions produced by electrons can be reconstructed as jets, leading to the same physical object being recorded as both an electron and a jet. To avoid using such jets in the analysis, which are actually electrons, jets which overlap within $\Delta R < 0.4$ with a electron satisfying the electron quality requirements detailed in Section 4.2.1 are removed. Likewise, any jet within $\Delta R < 0.4$ of a muon satisfying the requirements outlined in Section 4.3.1 is removed. Priority is given to muons over electrons and as such any electron within $\Delta R < 0.2$ of a muon is removed.

4.7 Conclusions

In this chapter the performance of identifying the expected signal features in the ATLAS detector was investigated through comparisons between Monte Carlo truth level and reconstructed objects. In particular, the leptons in the both the VBF and GF signal were shown to have a reconstruction performance only slightly degraded compared to sources of very isolated leptons ($\text{Z} \to \mu\mu/ee$) and a much improved performance compared to the more non-isolated source of leptons $\text{t}\bar{\text{t}}$. Good agreement

was found between data and Monte Carlo for the basic lepton variables investigated. Jet reconstruction performance in the signal was shown to be adequate and comparable compared to the other samples considered. A worse agreement between data and Monte Carlo was observed for jet related variables, although, from the low p_{T} nature of these events, it is possible this discrepancy is due to pile-up not being taken into account in the Monte Carlo samples used. These findings suggest a robust search strategy may be developed for the signal considered.

Chapter 5

Event Selection

In this chapter the cut-based selection used for the search for the SM Higgs boson in the H $\rightarrow$ ZZ $\rightarrow$ $ll\nu\nu$ channel produced by VBF is detailed. First the event pre-selection cuts are discussed and motivated. Subsequently discriminating variables associated with the Higgs decay products are described and a baseline set of cuts on these variables discussed. This selection is guided by the work carried out in [40, 41], also used for validation purposes. Next the variables useful to discriminate between signal and background that are associated to the VBF remnants are described. In each case the results of the baseline selections are summarised. In addition the effect of pile-up on each variable of interest is investigated and comparisons between data and Monte Carlo are made in order to ensure the variables used are well understood. For samples with an insufficient number of Monte Carlo events an attempt is made to estimate the contribution of these backgrounds.

5.1 Event pre-selection

The following outlines the pre-selection cuts used. Half of the available Monte Carlo events is used so that cuts can be optimized on an independent sample, which is discussed in Chapter 6. Firstly, at least one primary vertex with at least three associated tracks ($p_{\mathrm{T}} > 150\,\mathrm{MeV}/c$) is required.

In order to be able to record signal events, this analysis relies on two single lepton triggers, designed to exploit the high p_{T} leptons in the signal. The triggers used in the data were already discussed in Section 3.5. In the Monte Carlo samples used in this study, not all triggers used on the data were available in the ATHENA release used. In the electron channel (Z $\rightarrow$ ee) the L1_EM14 trigger is used while in the muon channel (Z $\rightarrow$ $\mu\mu$) the Event Filter chain EF_mu10_MG is used (same triggers used in pile-up Monte Carlo). These require an electron with transverse energy greater than 14 GeV and a muon with a transverse momentum of 10 GeV/c

	Trigger	Before			After		
		ee$\nu\nu$	$\mu\mu\nu\nu$	ee/$\mu\mu\nu\nu$	ee$\nu\nu$	$\mu\mu\nu\nu$	ee/$\mu\mu\nu\nu$
VBF	L1_EM14	99.14	54.07	76.89	100.00	54.98	75.40
	EF_mu10_MG	0.35	95.36	36.08	0.54	100.00	54.88
	L1_EM14‖EF_mu10_MG	99.14	98.16	98.68	100.00	100.00	100.00
	L1_EM14&&EF_mu10_MG	0.35	52.27	24.84	0.54	54.98	30.28
		ee$\nu\nu$	$\mu\mu\nu\nu$	ee/$\mu\mu\nu\nu$	ee$\nu\nu$	$\mu\mu\nu\nu$	ee/$\mu\mu\nu\nu$
GF	L1_EM14	98.31	20.36	57.15	100.00	18.82	53.77
	EF_mu10_MG	0.12	92.57	35.24	0.00	100.00	56.95
	L1_EM14‖EF_mu10_MG	98.31	94.05	96.23	100.00	100.00	100.00
	L1_EM14&&EF_mu10_MG	0.12	18.88	9.29	0.00	18.82	10.72

Table 5.1: Trigger selection efficiencies (%) for different combinations of electron (L1_EM14) and muon (EF_mu10_MG) triggers for VBF (upper section) and GF only (lower section) components of the $m_H = 200\,\mathrm{GeV}/c^2$ $H \to ZZ \to ll\nu\nu$ signal sample. Efficiencies are given for both the generated sample and the sub-sample containing two reconstructed electrons or muons passing the selection requirements outlined in Chapter 4. In each case this is broken down into H→ee$\nu\nu$,$\mu\mu\nu\nu$,ee+$\mu\mu\nu\nu$ events using truth information.

respectively. These triggers were shown to have comparable efficiency to those used in the data [40].

A comparison of the trigger efficiency of the VBF component of the signal compared to the GF signal, using the $m_H = 200\,\mathrm{GeV}/c^2$ signal sample, for each combination of triggers used on the Monte Carlo before and after the offline selection of the lepton candidates (described in Section 5.2) is shown in Table 5.1. The performance of the individual triggers is shown to be above 90 % in both the electron and muon channels for each signal component before the lepton selection. After the selection of lepton candidates this rises to 100 % in each case. This indicates that the use of single lepton triggers provides adequate performance for triggering both components of the signal. A certain level of overlap is shown to exist between triggers and channels although this appears to be much reduced with the muon trigger, where the muon trigger fires in the electron channel typically less than 1 % of the time. In contrast, for the electron trigger this value is closer to 50 %, caused by the looser nature of the electron trigger employed.

The effect of the pre-selection cuts is shown in Tables 5.2. As expected the vertex cut has a small but similar effect on all samples, typically reducing sample statistics by less than 1 %. For all samples considered the electron trigger is more efficient than the muon trigger. This is expected because the electron trigger is a Level 1 trigger whereas the muon trigger is an Event Filter trigger and therefore places much stricter requirements on the muons being triggered. Typically, the trigger efficiency for samples with isolated electrons(muons) is around 60(40) % respectively. For samples with non-isolated leptons, the triggering efficiencies are much lower at less than 30(20) % for electron(muons).

(a)

Sample	No Cuts	Vertex	Trigger	
			$Z \to ee$	$Z \to \mu\mu$
$m_H = 200\,GeV/c^2$				
VBF	3259	3256	2506	1176
GF	11607	11603	6633	4089
$m_H = 300\,GeV/c^2$				
VBF	3377	3364	2686	1263
GF	11481	11480	7582	4187
top	521863	521652	456418	190322
WW	219318	218823	128555	92910
ZZ	160967	160798	103395	75452
WZ	223381	223238	153074	102766
Z+jets	12689896	12661630	4752153	3792136
W+jets	2834715	2822945	838752	597406
Wbb/c+jets	654651	654447	270632	160871
Zbb+jets	447021	446898	230272	154144
QCD	12060648	12050134	6650166	2715590

(b)

Sample	No Cuts	Vertex	Trigger	
			$Z \to ee$	$Z \to \mu\mu$
$m_H = 200\,GeV/c^2$				
VBF	0.22±0.0039	0.22±0.0039	0.17±0.0034 (0.77)	0.08±0.0023 (0.36)
GF	1.81±0.02	1.81±0.02	1.04±0.01 (0.57)	0.64±0.01 (0.35)
$m_H = 300\,GeV/c^2$				
VBF	0.12±0.0022	0.12±0.0021	0.10±0.0019 (0.80)	0.05±0.001 (0.37)
GF	1.00±0.009	1.00±0.009	0.66±0.0076 (0.66)	0.37±0.0056 (0.36)
top	3503±5	3502±5	3022±5 (0.86)	1243±3 (0.35)
WW	163.8±0.4	163.4±0.4	94.99±0.27 (0.58)	68.98±0.23 (0.42)
ZZ	39.11±0.20	39.08±0.20	25.08±0.16 (0.64)	14.47±0.13 (0.37)
WZ	134.8±0.6	134.7±0.6	89.52±0.49 (0.66)	51.69±0.35 (0.38)
Z+jets	103516±29	103285±29	38969±18 (0.38)	30800±16 (0.30)
W+jets	684169±423	681491±422	201314±232 (0.29)	144058±194 (0.21)
Wbb/c+jets	26645±33	26637±33	11006±21 (0.41)	6547±16 (0.25)
Zbb+jets	1245±2	1244±2	612.4±1.3 (0.49)	424.6±1.1 (0.34)
QCD	$3.53\times10^{11}\pm3.95\times10^{8}$	$3.53\times10^{11}\pm3.94\times10^{8}$	$5.45\times10^{8}\pm4.62\times10^{6}$ (1.5×10^{-3})	$2.55\times10^{7}\pm1.81\times10^{6}$ (7.2×10^{-5})
Total	$3.53\times10^{11}\pm3.95\times10^{8}$	$3.53\times10^{11}\pm3.94\times10^{8}$	$5.5\times10^{8}\pm4.62\times10^{6}$ (1.5×10^{-3})	$2.6e\times10^{7}\pm1.81\times10^{6}$ (7.3×10^{-5})

Table 5.2: Expected number of (a) simulated events and (b) events normalised to the luminosity analysed in the data ($33.4\,pb^{-1}$) passing the pre-selection cuts outlined in Section 5.1. A break down of the $m_H = 200{,}300\,GeV/c^2$ signal samples is given in terms of the VBF and GF components. Where appropriate efficiencies are given in parentheses.

Primary vertex multiplicity	Data	MC	reweight(Data/MC)
1	0.237	0.183	1.297
2	0.322	0.279	1.156
3	0.242	0.265	0.911
4	0.123	0.164	0.746
5	0.053	0.074	0.719
6	0.017	0.025	0.669
7	0.005	0.007	0.670
8	0.001	0.002	0.583

Table 5.3: Pile-up reweight event weights applied to pile-up Monte Carlo as a function of event Primary vertex multiplicity.

5.2 Selection based on Higgs decay products

In the following the variables that provide good discrimination between signal and background that are related to the decay products of the Higgs boson are motivated and a baseline selection adopted. This baseline selection is based on the work performed recently in an analysis targeting the GF component of the signal [40, 41], which was also used for validation purposes. Cut variables identified in it were investigated in the context of the VBF signal and cuts changed appropriately. Comparisons are made with data and pile-up Monte Carlo reweighted to show similar levels of pile-up as that in the data analysed. By this procedure any regions where there are discrepancies between the nominal non-pile-up Monte Carlo and/or the data and pile-up reweighted Monte Carlo are avoided in order to try to maximise robustness of the analysis.

In order to correct the pile-up Monte Carlo to reflect the level of pile-up seen in the data an event reweighting procedure is performed on the pile-up Monte Carlo samples. In this procedure, the number of primary vertices in the event with three or more tracks is plotted after the di-lepton mass window cut (which is motivated later in this section) for the total background using the pile-up Monte Carlo and the data [40]. With each distribution normalised to unity the event weights as a function of the number of primary vertices shown in Table 5.3 were derived by taking the ratio of data over Monte Carlo for each bin. The primary vertex multiplicity for the total Monte Carlo and the data analysed before and after this reweighting procedure are shown in Figure 5.1. A much improved agreement between data and pile-up Monte Carlo is achieved when this Monte Carlo has been reweighted.

Lepton Selection In the following the choice of the selection of leptons is motivated.

Lepton p_{T} Typically backgrounds with no real leptons coming from the hard process will produce fake leptons which have a softer p_{T} distribution. For back-

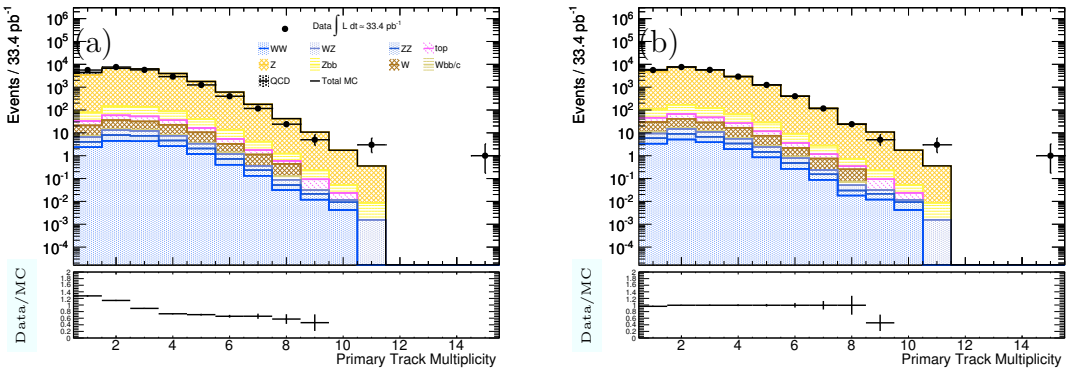


Figure 5.1: Distributions of primary track multiplicity for events with three or more associated tracks for pileup Monte Carlo (a) before and (b) after the pile-up reweighting procedure.

grounds with real leptons the p_{T} hardness is not much different to the signal. In this sense the lepton p_{T} cut serves to act to make sure the leptons selected are of good quality. This will allow further suppression of backgrounds without real leptons such as QCD. As detailed in Chapter 4, electrons used in this study are required to have $E_{\mathrm{T}} > 20\,\mathrm{GeV}$ and muons are required to have $p_{\mathrm{T}} > 20\,\mathrm{GeV}/c$. The p_{T} [1] distribution for the leading and sub-leading leptons passing the selection requirements are shown in Figures 5.2(a) and 5.2(b). The plots show contributions from each background considered and the sum of all Monte Carlo backgrounds, represented by the total MC line. The data analysed is included and is shown with the circular markers. The $\mathrm{m_H} = 200\,\mathrm{GeV}/c^2$ signal sample is drawn, with the VBF and GF components shown separately. Good agreement is found between data and total Monte Carlo background over the p_{T} range plotted for both the leading and sub-leading lepton within the statistical error as indicated by the Data/MC plot at the bottom of each histogram. Note that the same convention is used in all subsequent distributions presented in this chapter that show a breakdown of individual background contributions.

A comparison of the leading and sub-leading lepton p_{T} distributions for non pile-up and pile-up reweighted Monte Carlo are shown in Figure 5.2, where a comparison is made for the $\mathrm{m_H} = 200\,\mathrm{GeV}/c^2$ signal and the total background. As throughout this chapter, the pile-up Monte Carlo is presented by the filled distribution and the non pile-up by the unfilled distribution. Signal(total background) is shown in the red(blue) respectively. For both the leading and sub-leading lepton p_{T} distributions, there is no difference between the pile-up and non pile-up sample distributions in both the signal and the background. This is as expected as pile-up typically affects quantities relating to low p_{T} jets.

[1] Here p_{T} refers to electron cluster E_{T} and muon p_{T}. This was adopted for labelling clarity.

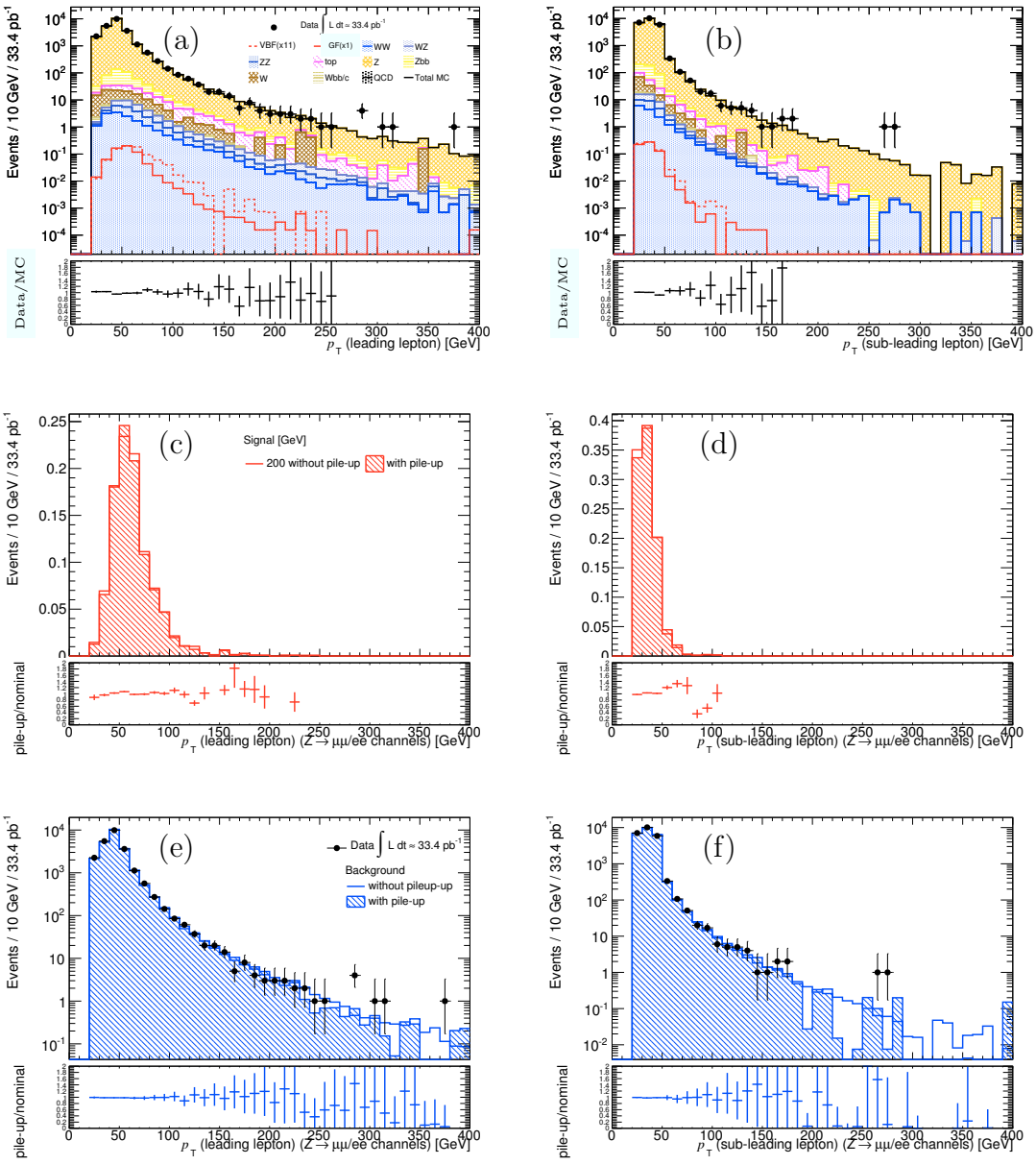


Figure 5.2: Distributions of the leading (left) and sub-leading (right) lepton p_{T}. The top-most plots (a,b) shows a breakdown of each background contribution and a comparison to data. The middle row of plots shows a comparison between pile-up reweighted and non pile-up Monte Carlo for the $\mathrm{m_H} = 200\,\mathrm{GeV}/c^2$ signal and the bottom row of plots shows a similar comparison for the total background. All the plots include contributions from the combined $\mathrm{Z} \to ee$ and $\mathrm{Z} \to \mu\mu$ channels.

Number of Leptons The presence of two leptons from the leptonically decaying Z boson in the signal topology allows for efficient rejection of backgrounds without isolated leptons, namely QCD. The distributions of the electron and muon multiplicities for events with either two electrons (corresponding to $Z \to ee$ in signal) or two muons ($Z \to \mu\mu$ in signal) satisfying the criteria outlined in Chapter 4 are shown in Figure 5.3. By requiring exactly two electrons or two muons and no leptons of any other type allows for rejection of some background, particularly the di-boson background. From now on $Z \to ee$ will be used to refer to events with exactly two electrons and no muons and $Z \to \mu\mu$ to events with exactly two muons. In the case of $Z \to \mu\mu$ in order to maintain good muon identification, at least one muon is required to be combined. It is possible the contamination of electrons in the $Z \to \mu\mu$ channel in the VBF signal is due to mis-identified electrons.

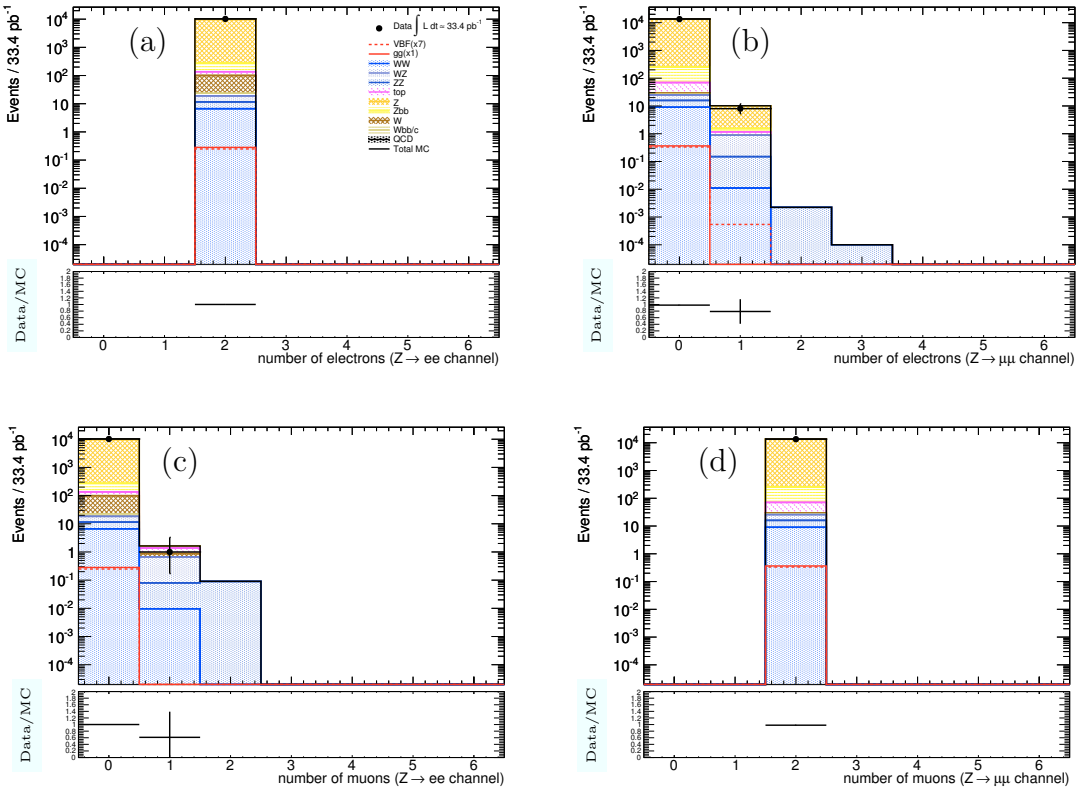


Figure 5.3: Distribution of electron and muon multiplicity in events with two electrons or two muons. Signal corresponds to the $m_H = 200\,\mathrm{GeV}/c^2$ sample and the contribution from each background (stacked) is shown.

Because the leptons in the signal are from a decaying Z, they are of the same flavour and have opposite electric charge. However, inefficiency in the measurement of electrons means that approximately 5 % of signal events with two electrons do not have opposite charge. This is represented in Figure 5.4, in which distributions of the product of selected lepton charges in (a) the $Z \to ee$ and (b) the $Z \to \mu\mu$ channels

respectively are shown. If the leptons are of opposite charge, the product of their charges is -1, otherwise it is +1. The mis-measurement of electron charge is seen in the signal by the overflow into the +1 bin. In order to retain as much signal as possible, only in the muon channel is the requirement of opposite sign leptons made, as in [40]. To summarise the lepton selection requires exactly two electrons or muons satisfying the respective lepton selection. Selected muons are required to have opposite electric charge. All plots presented in the remainder of this section have been made requiring this lepton selection.

The effect of the lepton selection and the subsequent cuts outlined in the following are summarised in Tables 5.4 and 5.5 showing the number of simulated events and the number of normalised events passing the cuts prescribed. After applying the lepton selection just detailed, an already very large suppression of QCD backgrounds is achieved, with its efficiency reduced to 2×10^{-9}. In the same way the W+jets background is strongly suppressed, now with an efficiency of 1×10^{-4}. By far the dominant background at this stage is Z+jets due to its very large cross section and high survival rate of the lepton selection. It makes up over 90 % of the total background. It is interesting to note that the $Z \to \tau\tau$+jets contribution is almost negligible compared to the contribution from $Z \to \mu\mu/ee$+jets. The efficiency of the GF and VBF components of the signal show similar performance at this stage of the selection, with approximately 35 % of signal being retained ($m_H = 200, 300\,\mathrm{GeV}/c^2$).

The performance of the selection appears to be similar in both the electron and muon channels, a trend which is apparent for all backgrounds with two leptons in the hard process. A different situation is seen for W+jets and QCD samples and typically samples which are not expected to have two real leptons. It is possible this is due to higher fake rates for electrons when compared to muons. Overall, about 30 % more background events are found in the muon channel as opposed to the electron channel. Fairly good agreement is seen between data and Monte Carlo. Worse agreement is seen in the electron channel. It is plausible this may be due to the use of crack electrons which in Section 4.2 were shown to be less well modelled in the Monte Carlo.

Di-lepton invariant mass The decay of one of the Higgs Z's into leptons implies that if the two leptons are reconstructed their mass should be close to the nominal Z mass. This characteristic of the signal is exploited by requiring the reconstructed di-lepton mass (m_{ll}) to be within a window of the nominal Z mass. An initial cut of $|m_{ll} - m_Z| < 20\,\mathrm{GeV}/c^2$ is used, where $m_Z = 91.1876\,\mathrm{GeV}/c^2$ is the nominal Z mass [1]. The di-lepton mass for combined $Z \to \mu\mu/ee$ channels is shown in Figure 5.5. A slight deficit is seen when data and Monte Carlo are compared.

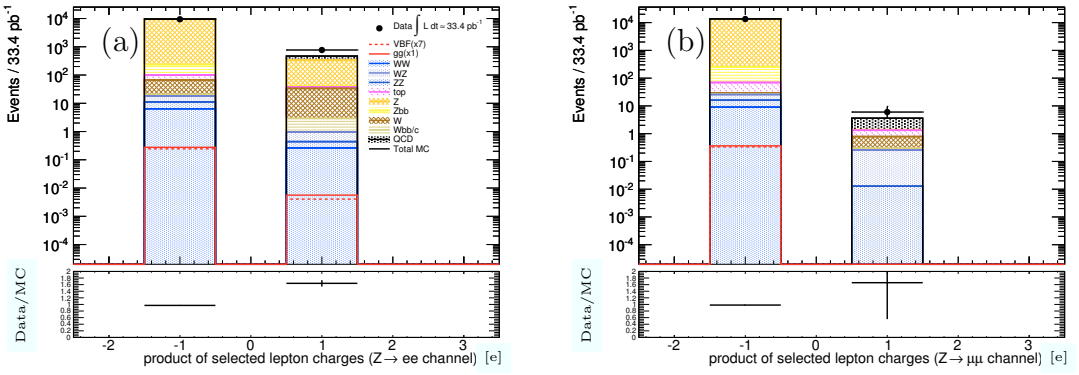


Figure 5.4: Distribution of products of lepton charges for (a) $Z \to ee$ and (b) $Z \to \mu\mu$ channels for the $m_H = 200\,\mathrm{GeV}/c^2$ sample and backgrounds (stacked).

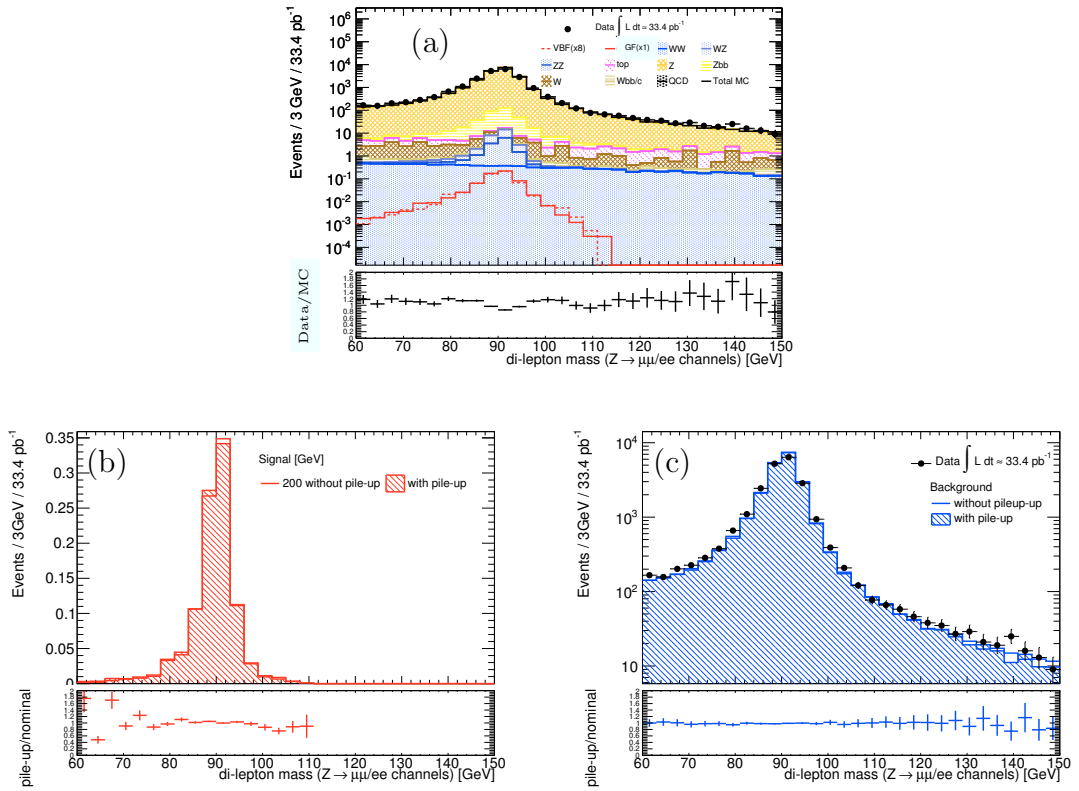


Figure 5.5: Distribution of di-lepton mass for (a) the $m_H = 200\,\mathrm{GeV}/c^2$ sample and backgrounds (stacked) and showing comparison of effect of pile-up reweighted Monte Carlo on (b) $m_H = 200\,\mathrm{GeV}/c^2$ signal and (c) total background.

The di-lepton mass cut is particularly efficient at rejecting backgrounds without decaying Z's and as such the backgrounds WW and W+jets are both reduced by approximately 70 %. Backgrounds with a top quark decay are also reduced by a similar factor. QCD is further suppressed by approximately 80 %. As expected the di-lepton mass window cut does not serve further to reduce backgrounds containing a real Z and as such Z+jets remains by far the largest background still making up over 90 % of the total background.

Missing Transverse Energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) The presence of two neutrinos from the decay of one of the Z bosons in the signal topology means that a significant amount of missing energy is present. In contrast backgrounds without any neutrinos in the final state should not have any associated missing energy, apart from detector inefficiencies. Therefore a cut on missing energy should be very efficient at rejecting backgrounds such as $\mathrm{Z} \rightarrow ll$+jets and QCD. As the Higgs boson becomes more massive its decay products become more and more boosted, giving rise to more $E_{\mathrm{T}}^{\mathrm{miss}}$. This behaviour can be seen in Figure 5.6(a) where the $E_{\mathrm{T}}^{\mathrm{miss}}$ in the signal samples as a function of Higgs mass is shown. In order to take advantage of this effect, at this stage in the analysis a low and high mass selection is proposed. The low mass selection is applied to signal masses $\mathrm{m_H} < 300\,\mathrm{GeV}/c^2$. In this region, $E_{\mathrm{T}}^{\mathrm{miss}}$ does not provide much discrimination between signal and background. A comparison between data and Monte Carlo for the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution is shown in Figure 5.6(b). A clear discrepancy between data and Monte Carlo is seen for $E_{\mathrm{T}}^{\mathrm{miss}} <$ 40 GeV, where the Monte Carlo is seen to underestimate the data. This is shown in figure 5.6(c) to be a result of pile-up, in which a much improved agreement to the data is seen by using the pile-up reweighted Monte Carlo. The effect of pile-up on the $\mathrm{m_H} = 200{,}300\,\mathrm{GeV}/c^2$ signal samples is negligible, as shown in Figure 5.6(d). This disagreement is understood as the low p_{T} jets from pile-up in the data contributing to higher levels of low $E_{\mathrm{T}}^{\mathrm{miss}}$ events. To avoid using this regime, for the low mass selection a baseline $E_{\mathrm{T}}^{\mathrm{miss}} > 40\,\mathrm{GeV}$ cut is applied. The high mass selection is applied for $\mathrm{m_H} \geq 300\,\mathrm{GeV}/c^2$, where the signal has enough $E_{\mathrm{T}}^{\mathrm{miss}}$ to provide substantial discrimination against background. For the high mass selection a baseline cut of $E_{\mathrm{T}}^{\mathrm{miss}} > 55\,\mathrm{GeV}$ is used.

Results shown in Tables 5.4-5.5 correspond to the use of the $\mathrm{m_H} = 200(300)\,\mathrm{GeV}/c^2$ Higgs samples in the low(high) mass selections respectively. As expected due to their intrinsic lack of missing energy, the missing energy cut strongly suppresses Z+jets background and further suppresses QCD, which are both reduced by over 90 % in the low mass selection. At this stage in the analysis, while Z+jets remains the largest background, the contribution from top becomes significant compared to Z+jets, having only been reduced by a comparable 25 % with the low mass selection $E_{\mathrm{T}}^{\mathrm{miss}}$ cut.

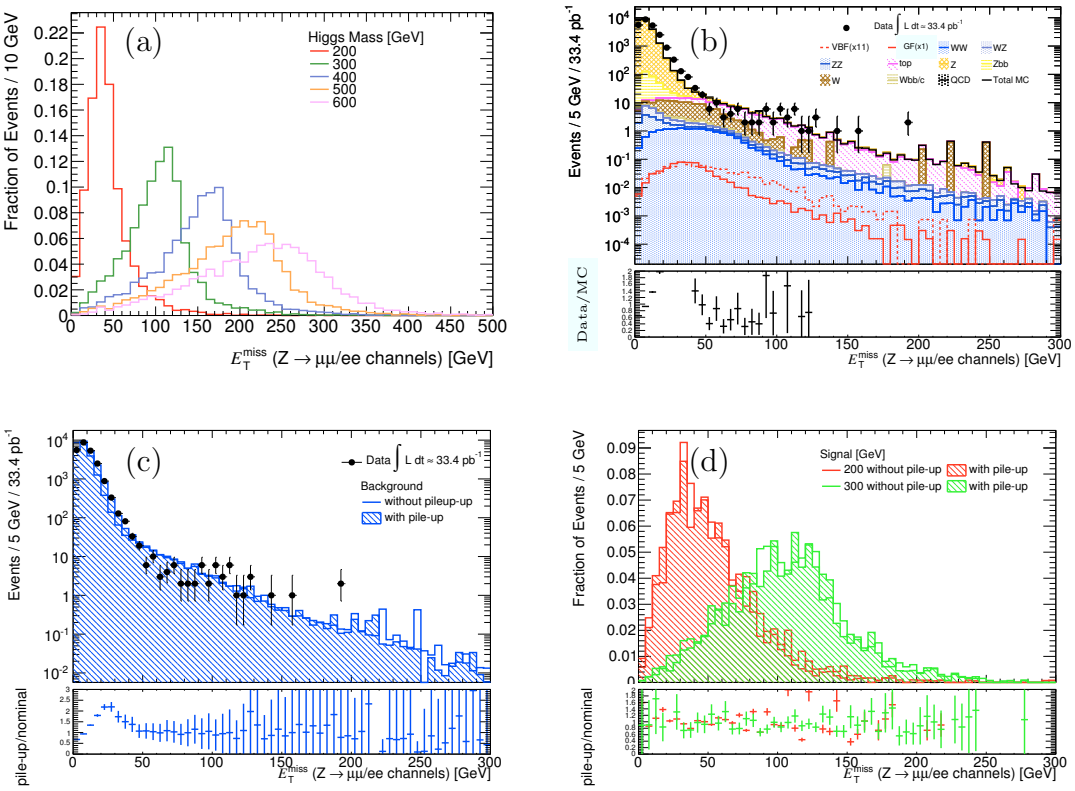


Figure 5.6: Distribution of $E_{\mathrm{T}}^{\mathrm{miss}}$ mass for (a) m_H = 200-600 GeV/c^2 signal samples in 100 GeV/c^2 steps, (b) the m_H = 200 GeV/c^2 sample and backgrounds (stacked). The effect of pile-up on the distribution is shown in (c) for the total background and in (d) for the m_H = 200,300 GeV/c^2 signals.

It is interesting to note that the contribution of single top is only a few percent of the total top background. It is strongly suppressed with the lepton selection and the di-lepton mass cuts because of the lack of two real leptons (especially in t channel production which makes up a large fraction of the single top contribution).

The $E_{\mathrm{T}}^{\mathrm{miss}}$ cut in the low mass selection also reduces the signal by up to 40 % in the case of the $\mathrm{m_H} = 200\,\mathrm{GeV}/c^2$ sample. However, this is by far outweighed by the very large suppression of the total background, giving rise to an improvement in signal significance (defined as $s/\sqrt{b}$, where s is signal and b is total background normalised to luminosity analysed in the data) relative to that after the di-lepton mass window cut of approximately a factor of 10, which is largely attributed to the strong suppression of the Z+jets background.

$\Delta\phi$ between selected leptons A further consequence of the Higgs decay products becoming more boosted with increasing Higgs mass is that the difference in ϕ between the leptons coming from the Z decay is reduced. This trend is shown in 5.7(a), which shows the $\Delta\phi$ between the identified leptons as a function of Higgs mass. Figures 5.7(b),5.7(c),5.7(d) show agreement between data and Monte Carlo for this variable is fairly good and the effect of pile-up on the signal and total background is negligible. For the low mass selection $\Delta\phi > 0.5$ is required and in the high mass selection $\Delta\phi < 2.0$ is used.

The $\Delta\phi$ cut appears to have similar impact on most of the backgrounds in the low mass selection, decreasing total background by approximately 10 %, while there is little expected change in the signal level. The effect of the cut is more prominent in the high mass selection, decreasing total background by 50 %. The most affected backgrounds are those with a Z, reduced by approximately 50 % each. In the high mass selection the main background now becomes $\mathrm{t\bar{t}}$, which makes up 40 % of the total background. In the low mass selection Z+jets remains the dominant background.

b-jet veto As discussed in Chapter 4 b-tagging should provide efficient rejection of backgrounds with b hadron decays. Events are rejected if they have any jet identified using the selection criteria detailed in Section 4.4.1 with $p_{\mathrm{T}} > 25\,\mathrm{GeV}/c^2$ and $|\eta| < 2.5$ and with a SV0 b-tagging weight > 5.72. Figure 5.8(a) shows the highest SV0 b-tagging weight of all jets satisfying the standard jet selection, for events passing the lepton selection. It clearly shows that the largest contribution of background satisfying the b-jet veto requirement comes from top quark backgrounds. This is further verified in Figure 5.8(b) where a comparison of the efficiency of the b-jet veto as a function of the jet p_{T} is made for the $\mathrm{m_H} = 200\,\mathrm{GeV}/c^2$ signal and the $\mathrm{t\bar{t}}$ background. In the signal an efficiency above 90 % is maintained across the entire

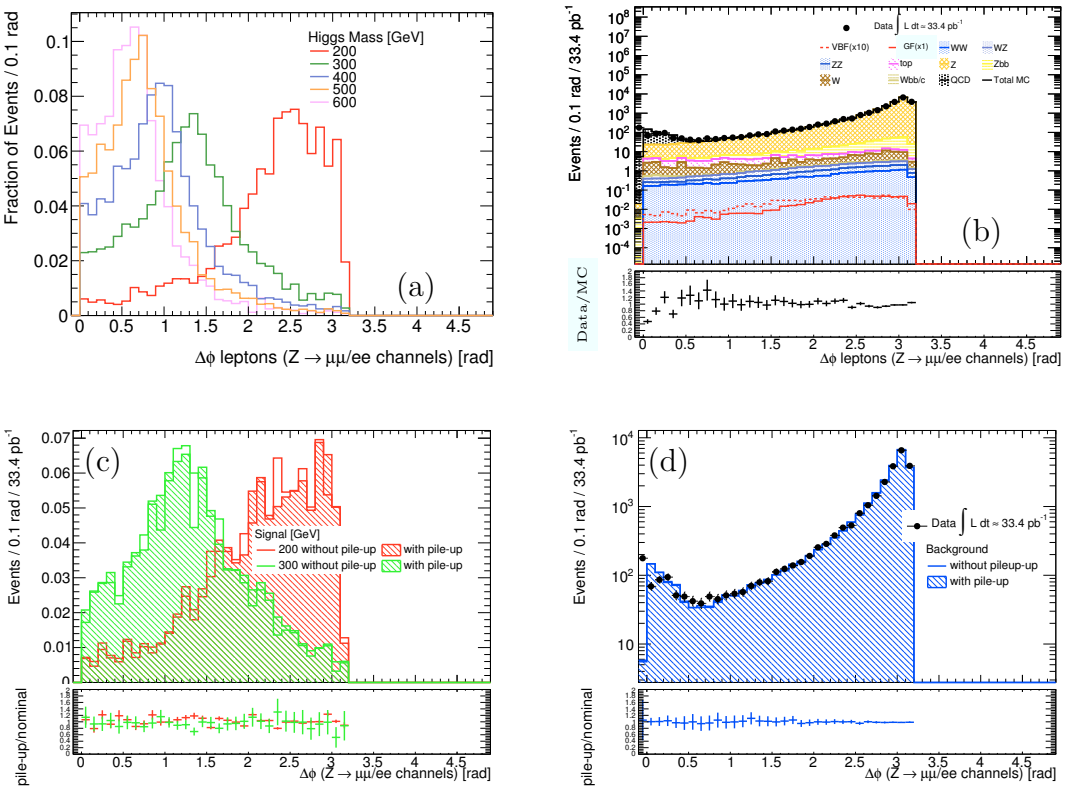


Figure 5.7: Distribution of $\Delta\phi$ between selected reconstructed leptons for (a) each signal sample with m_H = 200-600 GeV/c^2 in steps of 100 GeV/c^2, (b) all backgrounds stacked with data and m_H = 200 GeV/c^2 signal sample overlaid. A comparison of pile-up reweighted and non pile-up Monte Carlo is shown in (c) for the m_H = 200,300 GeV/c^2 signal and (d) the total background.

p_{T} range whereas in the $t\bar{t}$ sample the efficiency reduces with jet p_{T} from around 50-60 % for jet $50 < p_{\mathrm{T}} < 100\,\mathrm{GeV}/c$. A fluctuation in efficiency appears for jet $p_{\mathrm{T}} > 200\,\mathrm{GeV}/c$ in the $t\bar{t}$ sample due to the low statistics of higher p_{T} jets passing the b-jet veto. Fairly good agreement between Data and Monte Carlo is seen. The effect of pile-up on the SV0 jet weight distribution for the signal (total background) is shown in Figure 5.8(c)(5.8(d)). In both cases the effect is found to be small.

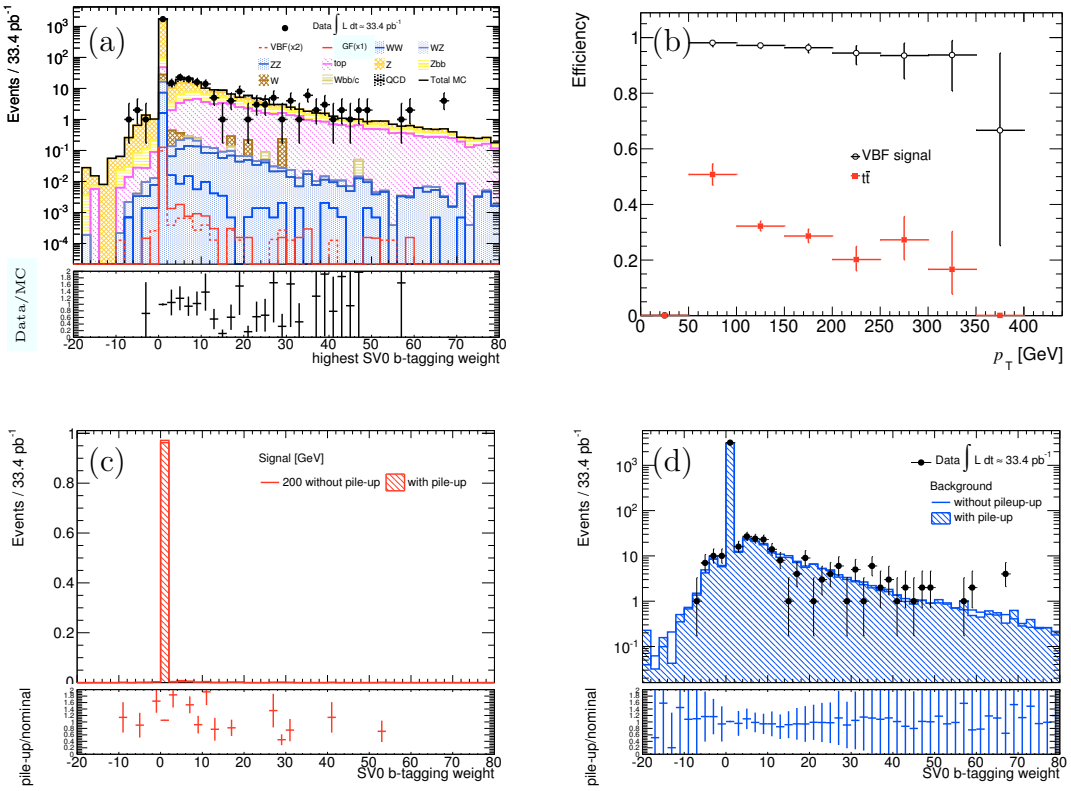


Figure 5.8: Distribution of SV0 b-tag weight for (a) highest weighted jet for signal and each background contribution (stacked), (b) the efficiency of the SV0 weight cut as a function of the leading jet p_{T} for the signal and $t\bar{t}$ and the effect of pile-up on the SV0 b-tagging weight of all jets for (c) the $\mathrm{m_H} = 200\,\mathrm{GeV}/c^2$ and (d) the total background.

The cut on SV0 tagging weight is negligible on all backgrounds other than those with b hadron decays, either from the decay of a top quark or b quarks in production with Z or W. These samples are suppressed by a further 30 % or more. An increase in significance of just less than 10 % is achieved in the low mass selection, similar to that seen in the high mass selection.

Transverse Mass (m_{T}) As there are two neutrinos in the signal final state the Higgs mass cannot be fully reconstructed. An approximation to this is to use the transverse mass m_{T}. The definition of transverse mass used in this analysis is

$$m_{\mathrm{T}} = \left(\sqrt{\mathrm{m_Z}^2 + |\vec{p_{\mathrm{T}}}(ll)|^2} + \sqrt{\mathrm{m_Z}^2 + |\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}|^2}\right)^2 - (\vec{p_{\mathrm{T}}}(ll) + \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}})^2 \qquad (5.1)$$

where $\mathrm{m_Z}$ is the nominal Z mass, $\vec{p_{\mathrm{T}}}(ll)$ refers to the p_{T} vector of the leptons and $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ the missing transverse momentum. Figure 5.9(a) shows how for the signal the peak of the m_{T} distribution falls approximately on the corresponding Higgs mass. From 5.9(b) the background distribution has a shape similar to the 200 GeV/c^2 signal and as such this is where discrimination is worst. As such no cut is placed on m_{T} in the low mass selection. With increasing Higgs mass better discrimination is achieved as more $E_{\mathrm{T}}^{\mathrm{miss}}$ is present in the signal. For the high mass selection a baseline cut of $m_{\mathrm{T}} > 200\,\mathrm{GeV}/c^2$ is applied.

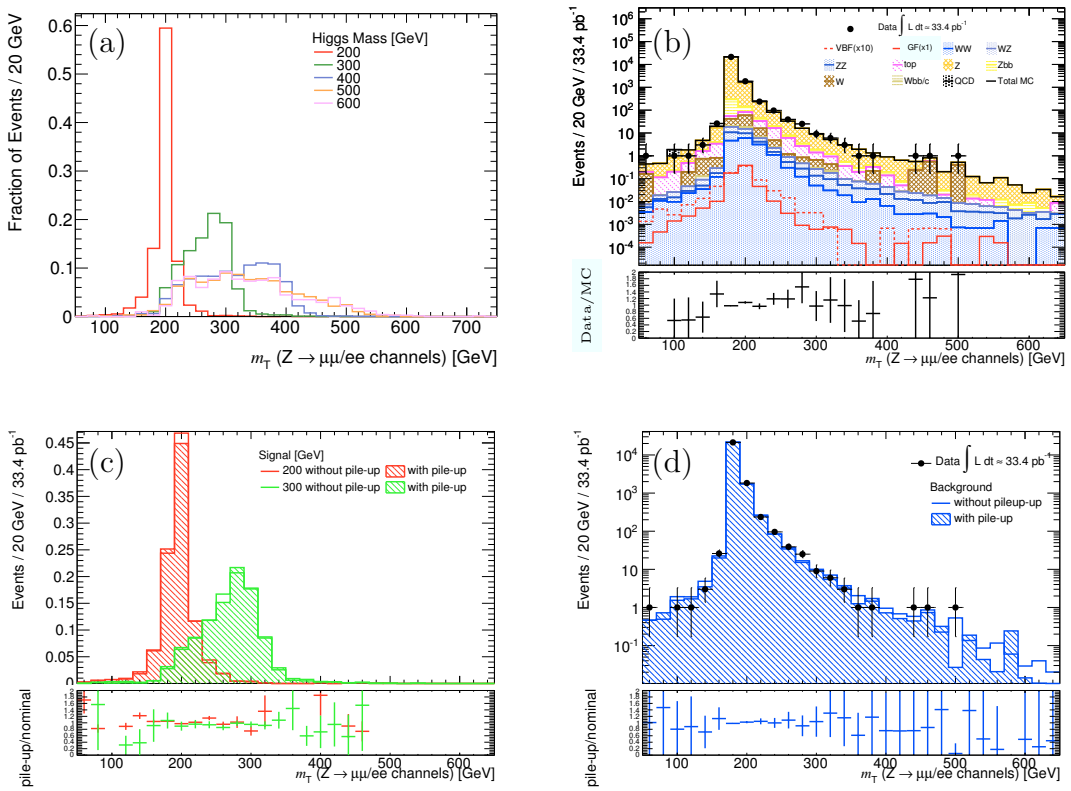


Figure 5.9: Distribution of m_{T} for (a) $\mathrm{m_H}$ = 200-600 GeV/c^2 signal samples in 100 GeV/c^2 steps, (b) the $\mathrm{m_H}$ = 200 GeV/c^2 sample and backgrounds (stacked). The effect of pile-up on the distribution is shown in (c) for the $\mathrm{m_H}$ = 200,300 GeV/c^2 signals and in (d) for the total background.

Good agreement is shown between data and Monte Carlo (5.9(b)) and the effect of pile-up on the signal (total background) is small 5.9(c)(5.9(d)).

Summary of Baseline Selection and Results

In the previous section a baseline analysis was presented for the cut-based selection of the VBF H $\rightarrow$ ZZ $\rightarrow$ $ll\nu\nu$ signal, relating to the Higgs decay products. In order to take advantage of the better discrimination between signal and background with increasing Higgs mass, a low and high mass selection were proposed. The selection is summarised in the following. Where there is a difference between low and high mass selections this is indicated with the low mass cut value defined and next to it the high mass cut value in brackets.

- Lepton Selection: Two electrons or muons passing requirements outlined in Sections 4.2.1 and 4.3.1. No other leptons of any other type. Opposite sign requirement made in muon channel.
- Di-lepton mass window: $|\mathrm{m}_{ll} - \mathrm{m}_{\mathrm{Z}}| < 20\,\mathrm{GeV}/c^2$.
- $E_{\mathrm{T}}^{\mathrm{miss}}$: $> 40(55)\,\mathrm{GeV}$.
- $\Delta\phi$: Between leptons $>0.5(<2.0)$ rad.
- b-jet veto: Reject any event with a jet with SV0 weight > 5.72.
- m_{T}: no cut$(> 200)\,\mathrm{GeV}/c^2$.

The number of simulated events and events normalised to the luminosity analysed in the data passing these selection cuts for the low and high mass selections are presented in Tables 5.4-5.6.

Baseline Low Mass Selection (Z → $\mu\mu/ee$ channels combined)						
Sample	lepton selection	di-lepton mass window	$E_{\mathrm{T}}^{\mathrm{miss}}$	$\Delta\phi$	b-jet veto	m_{T}
VBF	1142	1105	654	628	614	614
GF	4139	4024	1738	1702	1679	1679
$t\bar{t}$	11885	3435	2491	2345	811	811
WW	21921	6738	3058	2931	2900	2900
ZZ	38523	36336	14435	13618	13483	13483
WZ	50166	45566	10050	9477	9320	9320
Z+jets	2778532	2592931	2664	2297	2181	2181
W+jets	324	107	28	26	25	25
Wbb+jets	140	44	12	11	9	9
Zbb+jets	113884	107261	1401	1218	623	623
QCD	555	69	3	3	3	3
single top	120	31	19	19	10	10
Z→ $\tau\tau$+jets	7159	782	36	36	36	36

Baseline High Mass Selection (Z → $\mu\mu/ee$ channels combined)						
Sample	lepton selection	di-lepton mass window	$E_{\mathrm{T}}^{\mathrm{miss}}$	$\Delta\phi$	b-jet veto	m_{T}
VBF	1295	1233	1100	941	916	894
GF	4309	4095	3640	3212	3187	3133
$t\bar{t}$	11885	3435	1873	923	313	258
WW	21921	6738	1650	1013	1002	945
ZZ	38523	36336	9694	6502	6435	6302
WZ	50166	45566	6073	3765	3695	3530
Z+jets	2778532	2592931	748	519	476	444
W+jets	324	107	16	7	7	6
Wbb+jets	140	44	1	0	0	0
Zbb+jets	113884	107261	478	367	157	151
QCD	555	69	1	0	0	0
single top	120	31	12	5	1	1
Z→ $\tau\tau$+jets	7159	782	20	2	2	2

Table 5.4: Number of simulated events passing the baseline selection relating to Higgs decay products for the low and high mass selections (Section 5.2). Included are combined statistics for the Z → $\mu\mu/ee$ channels. The signal corresponds to the $m_{\mathrm{H}} = 200(300)\,\mathrm{GeV}/c^2$ sample for the low(high) mass selection.

Z → ee channel (Baseline Baseline Low Mass Selection)						
Sample	lepton selection	di-lepton mass window	$E_{\mathrm{T}}^{\mathrm{miss}}$	$\Delta\phi$	b-jet veto	m_{T}
VBF	0.04±0.002 (0.16)	0.03±0.002 (0.16)	0.02±0.001 (9.3×10^{-2})	0.02±0.001 (8.9×10^{-2})	0.02±0.001 (8.7×10^{-2})	0.02±0.001 (8.7×10^{-2})
GF	0.28±0.0066 (0.16)	0.27±0.0066 (0.15)	0.12±0.0043 (6.6×10^{-2})	0.12±0.0043 (6.5×10^{-2})	0.12±0.0043 (6.4×10^{-2})	0.12±0.0043 (6.4×10^{-2})
$t\bar{t}$	34.35±0.46 (1.1×10^{-2})	9.75±0.24 (3.2×10^{-3})	6.90±0.21 (2.3×10^{-3})	6.44±0.20 (2.1×10^{-3})	2.25±0.12 (7.4×10^{-4})	2.25±0.12 (7.4×10^{-4})
WW	6.67±0.07 (4.1×10^{-2})	2.08±0.04 (1.3×10^{-2})	0.92±0.03 (5.6×10^{-3})	0.88±0.03 (5.4×10^{-3})	0.87±0.03 (5.3×10^{-3})	0.87±0.03 (5.3×10^{-3})
ZZ	4.78±0.08 (0.12)	4.69±0.08 (0.12)	0.66±0.01 (1.7×10^{-2})	0.62±0.01 (1.6×10^{-2})	0.60±0.01 (1.5×10^{-2})	0.60±0.01 (1.5×10^{-2})
WZ	6.85±0.08 (5.1×10^{-2})	6.32±0.07 (4.7×10^{-2})	0.79±0.02 (5.9×10^{-3})	0.73±0.02 (5.4×10^{-3})	0.72±0.02 (5.4×10^{-3})	0.72±0.02 (5.4×10^{-3})
Z+jets	9385±9 (0.14)	8789±9 (0.13)	5.93±0.22 (8.6×10^{-5})	4.93±0.20 (7.1×10^{-5})	4.57±0.19 (6.6×10^{-5})	4.57±0.19 (6.6×10^{-5})
W+jets	76.75±4.94 (1.1×10^{-4})	26.14±3.05 (3.8×10^{-5})	6.66±1.96 (9.7×10^{-6})	6.11±1.96 (8.9×10^{-6})	5.96±1.92 (8.7×10^{-6})	5.96±1.92 (8.7×10^{-6})
Wbb+jets	4.42±0.47 (1.7×10^{-4})	1.26±0.33 (4.7×10^{-5})	0.34±0.24 (1.3×10^{-5})	0.34±0.24 (1.3×10^{-5})	0.26±0.20 (9.9×10^{-6})	0.26±0.20 (9.9×10^{-6})
Zbb+jets	138.4±0.7 (0.17)	130.7±0.6 (0.16)	1.33±0.06 (1.6×10^{-3})	1.14±0.06 (1.4×10^{-3})	0.55±0.04 (6.6×10^{-4})	0.55±0.04 (6.6×10^{-4})
QCD	923.8±327.2 (2.6×10^{-9})	178.3±237.9 (5.0×10^{-10})	2.12±2.88 (6.0×10^{-12})	2.12±2.88 (6.0×10^{-12})	2.12±2.88 (6.0×10^{-12})	2.12±2.88 (6.0×10^{-12})
single top	1.41±0.20 (3.2×10^{-3})	0.44±0.11 (9.8×10^{-4})	0.21±0.11 (4.6×10^{-4})	0.21±0.11 (4.6×10^{-4})	0.08±0.08 (1.9×10^{-4})	0.08±0.08 (1.9×10^{-4})
Z→ $\tau\tau$+jets	25.38±0.45 (7.3×10^{-4})	3.10±0.16 (8.9×10^{-5})	0.15±0.05 (4.3×10^{-6})	0.15±0.05 (4.3×10^{-6})	0.15±0.05 (4.3×10^{-6})	0.15±0.05 (4.3×10^{-6})
Total	10608±327 (3.0×10^{-8})	9152±238 (2.6×10^{-8})	26.0±3.5 (7.4×10^{-11})	23.7±3.5 (6.7×10^{-11})	18.1±3.5 (5.1×10^{-11})	18.1±3.5 (5.1×10^{-11})
Data	10199	8831	29	26	21	21

Z → $\mu\mu$ channel (Baseline Low Mass Selection)						
Sample	lepton selection	di-lepton mass window	$E_{\mathrm{T}}^{\mathrm{miss}}$	$\Delta\phi$	b-jet veto	m_{T}
VBF	0.04±0.002 (0.19)	0.04±0.002 (0.18)	0.02±0.001 (0.11)	0.02±0.001 (0.10)	0.02±0.001 (0.10)	0.02±0.001 (0.10)
GF	0.36±0.0075 (0.20)	0.35±0.0074 (0.20)	0.15±0.0049 (8.3×10^{-2})	0.15±0.0048 (8.2×10^{-2})	0.15±0.0048 (8.1×10^{-2})	0.15±0.0048 (8.1×10^{-2})
$t\bar{t}$	38.44±0.49 (1.3×10^{-2})	11.29±0.26 (3.7×10^{-3})	8.35±0.23 (2.7×10^{-3})	7.91±0.22 (2.6×10^{-3})	2.72±0.13 (8.9×10^{-4})	2.72±0.13 (8.9×10^{-4})
WW	8.98±0.08 (5.5×10^{-2})	2.73±0.04 (1.7×10^{-2})	1.27±0.03 (7.7×10^{-3})	1.22±0.03 (7.4×10^{-3})	1.20±0.03 (7.4×10^{-3})	1.20±0.03 (7.4×10^{-3})
ZZ	6.81±0.09 (0.17)	6.69±0.09 (0.17)	0.90±0.01 (2.3×10^{-2})	0.85±0.01 (2.2×10^{-2})	0.82±0.01 (2.1×10^{-2})	0.82±0.01 (2.1×10^{-2})
WZ	8.92±0.09 (6.6×10^{-2})	8.63±0.08 (6.4×10^{-2})	0.99±0.02 (7.4×10^{-3})	0.93±0.02 (6.9×10^{-3})	0.91±0.02 (6.8×10^{-3})	0.91±0.02 (6.8×10^{-3})
Z+jets	13310±10 (0.19)	12414±10 (0.18)	15.43±0.35 (2.2×10^{-4})	13.50±0.33 (2.0×10^{-4})	12.95±0.32 (1.9×10^{-4})	12.95±0.32 (1.9×10^{-4})
W+jets	2.28±1.42 (3.3×10^{-6})	1.39±1.24 (2.0×10^{-6})	0.16±0.37 (2.4×10^{-7})	0.16±0.37 (2.4×10^{-7})	0.16±0.37 (2.4×10^{-7})	0.16±0.37 (2.4×10^{-7})
Wbb+jets	1.26±0.30 (4.7×10^{-5})	0.52±0.25 (2.0×10^{-5})	0.12±0.14 (4.6×10^{-6})	0.08±0.13 (3.1×10^{-6})	0.08±0.13 (3.1×10^{-6})	0.08±0.13 (3.1×10^{-6})
Zbb+jets	181.7±0.7 (0.22)	171.3±0.7 (0.21)	2.02±0.07 (2.4×10^{-3})	1.78±0.07 (2.1×10^{-3})	0.95±0.05 (1.1×10^{-3})	0.95±0.05 (1.1×10^{-3})
QCD	114.9±11.7 (3.3×10^{-10})	1.11±2.54 (3.1×10^{-12})	-	-	-	-
single top	1.76±0.22 (3.9×10^{-3})	0.33±0.13 (7.4×10^{-4})	0.27±0.12 (6.1×10^{-4})	0.27±0.12 (6.1×10^{-4})	0.16±0.10 (3.5×10^{-4})	0.16±0.10 (3.5×10^{-4})
Z→ $\tau\tau$+jets	30.78±0.50 (8.8×10^{-4})	3.08±0.16 (8.8×10^{-5})	0.14±0.05 (4.0×10^{-6})	0.14±0.05 (4.0×10^{-6})	0.14±0.05 (4.0×10^{-6})	0.14±0.05 (4.0×10^{-6})
Total	13706±16 (3.9×10^{-8})	12621±11 (3.6×10^{-8})	29.7±0.6 (8.4×10^{-11})	26.8±0.6 (7.6×10^{-11})	20.1±0.5 (5.7×10^{-11})	20.1±0.5 (5.7×10^{-11})
Data	13434	12294	25	24	20	20

Z → $\mu\mu/ee$ channels combined (Baseline Low Mass Selection)						
Sample	lepton selection	di-lepton mass window	$E_{\mathrm{T}}^{\mathrm{miss}}$	$\Delta\phi$	b-jet veto	m_{T}
VBF	0.08±0.002 (0.35)	0.07±0.002 (0.34)	0.04±0.002 (0.20)	0.04±0.002 (0.19)	0.04±0.002 (0.19)	0.04±0.002 (0.19)
GF	0.65±0.01 (0.36)	0.63±0.0099 (0.35)	0.27±0.0065 (0.15)	0.27±0.0064 (0.15)	0.26±0.0064 (0.14)	0.26±0.0064 (0.14)
$t\bar{t}$	72.78±0.67 (2.4×10^{-2})	21.03±0.36 (6.9×10^{-3})	15.26±0.31 (5.0×10^{-3})	14.36±0.30 (4.7×10^{-3})	4.97±0.17 (1.6×10^{-3})	4.97±0.17 (1.6×10^{-3})
WW	15.65±0.11 (9.6×10^{-2})	4.81±0.06 (2.9×10^{-2})	2.19±0.04 (1.3×10^{-2})	2.10±0.04 (1.3×10^{-2})	2.07±0.04 (1.3×10^{-2})	2.07±0.04 (1.3×10^{-2})
ZZ	11.59±0.12 (0.30)	11.38±0.12 (0.29)	1.56±0.02 (4.0×10^{-2})	1.46±0.02 (3.7×10^{-2})	1.42±0.02 (3.6×10^{-2})	1.42±0.02 (3.6×10^{-2})
WZ	15.78±0.11 (0.12)	14.96±0.11 (0.11)	1.79±0.03 (1.3×10^{-2})	1.67±0.03 (1.2×10^{-2})	1.63±0.03 (1.2×10^{-2})	1.63±0.03 (1.2×10^{-2})
Z+jets	22695±14 (0.33)	21203±13 (0.31)	21.36±0.41 (3.1×10^{-4})	18.44±0.39 (2.7×10^{-4})	17.51±0.38 (2.5×10^{-4})	17.51±0.38 (2.5×10^{-4})
W+jets	79.03±5.14 (1.2×10^{-4})	27.53±3.29 (4.0×10^{-5})	6.82±2.00 (1.0×10^{-5})	6.27±1.99 (9.2×10^{-6})	6.12±1.96 (8.9×10^{-6})	6.12±1.96 (8.9×10^{-6})
Wbb+jets	5.68±0.52 (2.1×10^{-4})	1.78±0.34 (6.7×10^{-5})	0.46±0.25 (1.7×10^{-5})	0.42±0.25 (1.6×10^{-5})	0.35±0.21 (1.3×10^{-5})	0.35±0.21 (1.3×10^{-5})
Zbb+jets	320.1±1 (0.39)	302±1 (0.36)	3.35±0.09 (4.0×10^{-3})	2.92±0.09 (3.5×10^{-3})	1.50±0.06 (1.8×10^{-3})	1.50±0.06 (1.8×10^{-3})
QCD	1039±327 (2.9×10^{-9})	179.5±237.9 (5.1×10^{-10})	2.12±2.88 (6.0×10^{-12})	2.12±2.88 (6.0×10^{-12})	2.12±2.88 (6.0×10^{-12})	2.12±2.88 (6.0×10^{-12})
single top	3.16±0.30 (7.1×10^{-3})	0.77±0.15 (1.7×10^{-3})	0.48±0.12 (1.1×10^{-3})	0.48±0.12 (1.1×10^{-3})	0.24±0.12 (5.4×10^{-4})	0.24±0.12 (5.4×10^{-4})
Z→ $\tau\tau$+jets	56.16±0.67 (1.6×10^{-3})	6.18±0.22 (1.8×10^{-4})	0.29±0.06 (8.2×10^{-6})	0.29±0.06 (8.2×10^{-6})	0.29±0.06 (8.2×10^{-6})	0.29±0.06 (8.2×10^{-6})
Total	24314±328 (6.9×10^{-8})	21773±238 (6.2×10^{-8})	55.66±3.56 (1.6×10^{-10})	50.51±3.55 (1.4×10^{-10})	38.22±3.52 (1.1×10^{-10})	38.22±3.52 (1.1×10^{-10})
Data	23633	21125	54	50	41	41
$s/\sqrt{b}$	0.0005±0.00001	0.0005±0.00005	0.006±0.0002	0.006±0.0002	0.007±0.0003	0.007±0.0003

Table 5.5: Expected number of simulated events passing H → ZZ → $ll\nu\nu$ cuts using low mass selection given an integrated luminosity of $33.4\,\mathrm{pb}^{-1}$, for Z → ee, Z → $\mu\mu$ and combined channel. Efficiencies are given in brackets. The signal corresponds to the $\mathrm{m_H} = 200\,\mathrm{GeV}/c^2$ sample.

Z → *ee* channel (Baseline High Mass Selection)						
Sample	lepton selection	di-lepton mass window	$E_{\mathrm{T}}^{\mathrm{miss}}$	$\Delta\phi$	b-jet veto	m_{T}
VBF	0.02±0.0009 (0.17)	0.02±0.0009 (0.16)	0.02±0.0008 (0.14)	0.02±0.0008 (0.13)	0.02±0.0008 (0.12)	0.01±0.0007 (0.12)
GF	0.17±0.0038 (0.17)	0.16±0.0037 (0.16)	0.14±0.0035 (0.14)	0.12±0.0033 (0.12)	0.12±0.0033 (0.12)	0.12±0.0032 (0.12)
$t\bar{t}$	34.35±0.46 (1.1×10^{-2})	9.75±0.24 (3.2×10^{-3})	5.06±0.18 (1.7×10^{-3})	2.37±0.12 (7.7×10^{-4})	0.82±0.07 (2.7×10^{-4})	0.67±0.06 (2.2×10^{-4})
WW	6.67±0.07 (4.1×10^{-2})	2.08±0.04 (1.3×10^{-2})	0.51±0.02 (3.1×10^{-3})	0.30±0.02 (1.9×10^{-3})	0.30±0.01 (1.8×10^{-3})	0.28±0.01 (1.7×10^{-3})
ZZ	4.78±0.08 (0.12)	4.69±0.08 (0.12)	0.44±0.0093 (1.1×10^{-2})	0.30±0.0085 (7.8×10^{-3})	0.30±0.0077 (7.6×10^{-3})	0.29±0.0076 (7.4×10^{-3})
WZ	6.85±0.08 (5.1×10^{-2})	6.32±0.07 (4.7×10^{-2})	0.46±0.02 (3.4×10^{-3})	0.29±0.01 (2.2×10^{-3})	0.29±0.01 (2.1×10^{-3})	0.27±0.01 (2.0×10^{-3})
Z+jets	9385±9 (0.14)	8789±9 (0.13)	1.63±0.12 (2.4×10^{-5})	1.42±0.11 (2.1×10^{-5})	1.30±0.10 (1.9×10^{-5})	1.18±0.10 (1.7×10^{-5})
W+jets	76.75±4.94 (1.1×10^{-4})	26.14±3.05 (3.8×10^{-5})	3.40±1.57 (5.0×10^{-6})	1.63±1.26 (2.4×10^{-6})	1.63±1.26 (2.4×10^{-6})	1.63±1.26 (2.4×10^{-6})
Wbb+jets	4.42±0.47 (1.7×10^{-4})	1.26±0.33 (4.7×10^{-5})	0.04±0.09 (1.4×10^{-6})	-	-	-
Zbb+jets	138.4±0.7 (0.17)	130.7±0.7 (0.16)	0.42±0.03 (5.1×10^{-4})	0.34±0.03 (4.1×10^{-4})	0.14±0.02 (1.7×10^{-4})	0.14±0.02 (1.6×10^{-4})
QCD	923.8±327.2 (2.6×10^{-9})	178.3±237.9 (5.0×10^{-10})	1.04±2.51 (2.9×10^{-12})	-	-	-
single top	1.41±0.20 (3.2×10^{-3})	0.44±0.11 (9.8×10^{-4})	0.12±0.09 (2.7×10^{-4})	0.07±0.08 (1.5×10^{-4})	-	-
Z→ ττ+jets	25.38±0.45 (7.3×10^{-4})	3.10±0.16 (8.9×10^{-5})	0.06±0.04 (1.8×10^{-6})	0.002±0.005 (5.5×10^{-8})	0.002±0.005 (5.5×10^{-8})	0.002±0.005 (5.5×10^{-8})
Total	10608±327 (3.0×10^{-8})	9152±238 (2.6×10^{-8})	13.2±3.0 (3.7×10^{-11})	6.7±1.3 (1.9×10^{-11})	4.8±1.3 (1.4×10^{-11})	4.5±1.3 (1.3×10^{-11})
Data	10199	8831	11	6	5	4

Z → *μμ* channel (Baseline High Mass Selection)						
Sample	lepton selection	di-lepton mass window	$E_{\mathrm{T}}^{\mathrm{miss}}$	$\Delta\phi$	b-jet veto	m_{T}
VBF	0.03±0.001 (0.21)	0.03±0.001 (0.20)	0.02±0.0009 (0.18)	0.02±0.0008 (0.15)	0.02±0.0008 (0.15)	0.02±0.0008 (0.15)
GF	0.21±0.0043 (0.21)	0.20±0.0042 (0.20)	0.18±0.0039 (0.18)	0.16±0.0037 (0.16)	0.16±0.0037 (0.16)	0.15±0.0037 (0.15)
$t\bar{t}$	38.44±0.49 (1.3×10^{-2})	11.29±0.26 (3.7×10^{-3})	6.40±0.20 (2.1×10^{-3})	3.28±0.14 (1.1×10^{-3})	1.10±0.08 (3.6×10^{-4})	0.91±0.07 (3.0×10^{-4})
WW	8.98±0.08 (5.5×10^{-2})	2.73±0.04 (1.7×10^{-2})	0.67±0.02 (4.1×10^{-3})	0.42±0.02 (2.5×10^{-3})	0.41±0.02 (2.5×10^{-3})	0.39±0.02 (2.4×10^{-3})
ZZ	6.81±0.09 (0.17)	6.69±0.09 (0.17)	0.60±0.0098 (1.5×10^{-2})	0.40±0.0086 (1.0×10^{-2})	0.39±0.0089 (9.9×10^{-3})	0.38±0.0087 (9.7×10^{-3})
WZ	8.92±0.09 (6.6×10^{-2})	8.63±0.08 (6.4×10^{-2})	0.59±0.01 (4.4×10^{-3})	0.37±0.01 (2.8×10^{-3})	0.36±0.01 (2.7×10^{-3})	0.35±0.01 (2.6×10^{-3})
Z+jets	13310±10 (0.19)	12414±10 (0.18)	4.35±0.19 (6.3×10^{-5})	2.72±0.15 (3.9×10^{-5})	2.50±0.14 (3.6×10^{-5})	2.37±0.14 (3.4×10^{-5})
W+jets	2.28±1.42 (3.3×10^{-6})	1.39±1.24 (2.0×10^{-6})	0.16±0.37 (2.4×10^{-7})	0.16±0.37 (2.4×10^{-7})	0.16±0.37 (2.4×10^{-7})	-
Wbb+jets	1.26±0.30 (4.7×10^{-5})	0.52±0.25 (2.0×10^{-5})	-	-	-	-
Zbb+jets	181.7±0.7 (0.22)	171.3±0.7 (0.21)	0.70±0.04 (8.5×10^{-4})	0.54±0.04 (6.5×10^{-4})	0.23±0.02 (2.7×10^{-4})	0.22±0.02 (2.7×10^{-4})
QCD	114.9±11.7 (3.3×10^{-10})	1.11±2.54 (3.1×10^{-12})	-	-	-	-
single top	1.76±0.22 (3.9×10^{-3})	0.33±0.13 (7.4×10^{-4})	0.19±0.11 (4.2×10^{-4})	0.06±0.07 (1.3×10^{-4})	0.03±0.07 (6.4×10^{-5})	0.03±0.07 (6.4×10^{-5})
Z→ ττ+jets	30.78±0.50 (8.8×10^{-4})	3.08±0.16 (8.8×10^{-5})	0.09±0.05 (2.6×10^{-6})	0.008±0.02 (2.3×10^{-7})	0.008±0.02 (2.3×10^{-7})	0.008±0.02 (2.3×10^{-7})
Total	13706±16 (3.9×10^{-8})	12621±11 (3.6×10^{-8})	13.8±0.5 (3.9×10^{-11})	8.0±0.4 (2.3×10^{-11})	5.2±0.4 (1.5×10^{-11})	4.6±0.2 (1.3×10^{-11})
Data	13434	12294	8	5	5	5

Z → *μμ*/*ee* channels combined (Baseline High Mass Selection)						
Sample	lepton selection	di-lepton mass window	$E_{\mathrm{T}}^{\mathrm{miss}}$	$\Delta\phi$	b-jet veto	m_{T}
VBF	0.05±0.001 (0.38)	0.05±0.001 (0.37)	0.04±0.001 (0.33)	0.03±0.001 (0.28)	0.03±0.001 (0.27)	0.03±0.001 (0.26)
GF	0.38±0.0057 (0.38)	0.36±0.0056 (0.36)	0.32±0.0052 (0.32)	0.28±0.0049 (0.28)	0.28±0.0049 (0.28)	0.27±0.0049 (0.27)
$t\bar{t}$	72.78±0.67 (2.4×10^{-2})	21.03±0.36 (6.9×10^{-3})	11.47±0.27 (3.8×10^{-3})	5.65±0.19 (1.8×10^{-3})	1.92±0.11 (6.3×10^{-4})	1.58±0.10 (5.2×10^{-4})
WW	15.65±0.11 (9.6×10^{-2})	4.81±0.06 (2.9×10^{-2})	1.18±0.03 (7.2×10^{-3})	0.72±0.02 (4.4×10^{-3})	0.71±0.02 (4.4×10^{-3})	0.67±0.02 (4.1×10^{-3})
ZZ	11.59±0.12 (0.3)	11.38±0.12 (0.29)	1.03±0.01 (2.6×10^{-2})	0.70±0.01 (1.8×10^{-2})	0.68±0.01 (1.8×10^{-2})	0.67±0.01 (1.7×10^{-2})
WZ	15.78±0.11 (0.12)	14.96±0.11 (0.11)	1.05±0.02 (7.8×10^{-3})	0.66±0.01 (4.9×10^{-3})	0.65±0.01 (4.8×10^{-3})	0.62±0.01 (4.6×10^{-3})
Z+jets	22695±14 (0.33)	21203±13 (0.31)	5.98±0.22 (8.7×10^{-5})	4.14±0.18 (6.0×10^{-5})	3.80±0.17 (5.5×10^{-5})	3.55±0.17 (5.1×10^{-5})
W+jets	79.03±5.14 (1.2×10^{-4})	27.53±3.29 (4.0×10^{-5})	3.57±1.61 (5.2×10^{-6})	1.79±1.31 (2.6×10^{-6})	1.79±1.31 (2.6×10^{-6})	1.63±1.26 (2.4×10^{-6})
Wbb+jets	5.68±0.52 (2.1×10^{-4})	1.78±0.34 (6.7×10^{-5})	0.04±0.09 (1.4×10^{-6})	-	-	-
Zbb+jets	320.1±1 (0.39)	302±1 (0.36)	1.13±0.05 (1.4×10^{-3})	0.88±0.05 (1.1×10^{-3})	0.37±0.03 (4.5×10^{-4})	0.36±0.03 (4.3×10^{-4})
QCD	1039±327 (2.9×10^{-9})	179.5±237.9 (5.1×10^{-10})	1.04±2.51 (2.9×10^{-12})	-	-	-
single top	3.16±0.30 (7.1×10^{-3})	0.77±0.15 (1.7×10^{-3})	0.31±0.13 (6.9×10^{-4})	0.13±0.10 (2.8×10^{-4})	0.03±0.07 (6.4×10^{-5})	0.03±0.07 (6.4×10^{-5})
Z→ ττ+jets	56.16±0.67 (1.6×10^{-3})	6.18±0.22 (1.8×10^{-4})	0.15±0.05 (4.4×10^{-6})	0.01±0.02 (2.9×10^{-7})	0.01±0.02 (2.9×10^{-7})	0.01±0.02 (2.9×10^{-7})
Total	24314±328 (6.9×10^{-8})	21773±238 (6.2×10^{-8})	26.95±3.01 (7.6×10^{-11})	14.69±1.35 (4.2×10^{-11})	9.97±1.33 (2.8×10^{-11})	9.11±1.28 (2.6×10^{-11})
Data	23633	21125	19	11	10	9
$s/\sqrt{b}$	0.0003±0.000009	0.0003±0.000009	0.008±0.0003	0.009±0.0003	0.011±0.0004	0.011±0.0004

Table 5.6: Expected number of simulated events passing $\mathrm{H} \to \mathrm{ZZ} \to \mathit{llνν}$ cuts using high mass selection given an integrated luminosity of $33.4\,\mathrm{pb}^{-1}$, for $\mathrm{Z} \to \mathit{ee}$, $\mathrm{Z} \to \mu\mu$ and combined channel. Efficiencies are given in brackets. The signal corresponds to the $\mathrm{m_H} = 300\,\mathrm{GeV}/c^2$ sample.

5.3 Selection of remnants of VBF process

In the following the set of cuts used to exploit the VBF topology in the signal is discussed. This will be referred to as the tag-jet selection whereby tag-jets is used to refer to the jets identified as originating from the VBF process. The tag-jet selection begins by requiring at least two jets passing the jet selection ($N_{jets} \geq 2$) described in Section 4.4.1, because it is expected signal events produced via VBF will typically have two reconstructed tag-jets. This is verified in Figure 5.10 which shows the jet multiplicity of jets passing the jet selection. Clearly the requirement of two jets passing the jet selection provides very good rejection of backgrounds like Z+jets. The jets associated with the VBF remnants are assumed to be the two highest p_T jets in the event, which is one of the most common ways to identify them [83]. Other possible methods are explored in Chapter 6.

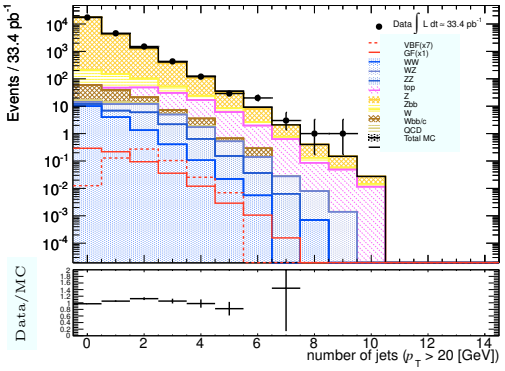


Figure 5.10: Multiplicity of jets passing selection requirements for signal (showing VBF and GF components separately) and background components added together.

Effect of overlap removal on tag-jets As discussed in Section 4.6, within this analysis jets overlapping geometrically with electrons or muons passing the quality requirements are removed from consideration. By far the dominant type of overlap is between electrons and jets. In the following study where the $m_H = 200\,\mathrm{GeV}/c^2$ signal sample was used, the total fraction of electron-jet overlaps out of all the overlaps was found to be over 99 %. It is important that the overlap removal does not remove the identified tag-jets. In the study performed the identified reconstructed tag-jets were matched to the truth quarks associated with the VBF topology (VBF quarks) using ΔR matching. The ΔR between each possibility of truth-reconstructed quark-tag-jet was calculated and the smallest summed total value then taken to establish which reconstructed jet is compared to which truth VBF quark. This was done separately for the $Z \to ee$ and $Z \to \mu\mu$ channels. The same procedure was then performed using the reconstructed jets removed in

the overlap removal procedure, to find those which most closely matched the truth VBF quarks. Figure 5.11(a) shows a comparison of the difference in p_{T} between each identified reconstructed tag-jet (p_{T}(reco)) and the corresponding matched truth VBF quark (p_{T}(truth)) as a function of the truth VBF quark p_{T}. It shows that in both the electron and muon channels there are reconstructed jets left after the overlap removal that far more accurately mimic the VBF quarks than those which are removed in the overlap removal process. This behaviour is seen across most of the p_{T} range investigated, indicating it is rare that the overlap removal process is rejecting jets which are very likely to be the identified tag-jets. The same behaviour is less apparent as a function of truth VBF quark η shown in Figure 5.11(b). However, this behaviour is further backed up in Figures 5.11(c) and 5.11(d), where it is shown there is a much improved agreement in terms the minimum ΔR, ΔR_{min}, between truth VBF quarks and the best matched jets not removed by the overlap procedure compared to those removed (Figure 5.11(c)) and this is verified if a comparison is made of the η distribution of the jets removed by the overlap removal and the truth VBF quarks (Figure 5.11(d)).

Tag-jet kinematics The tag-jets in the signal are derived from the hard process and result from recoiling against the Higgs boson. Because of this they tend to be harder than the jets found in the backgrounds. This behaviour is shown in Figure 5.12(a) where the p_{T} of the leading tag-jet is shown. It is plotted requiring the lepton selection detailed previously in this chapter as well as two or more jets passing the quality requirements detailed in Section 4.4.1. This is true of all subsequent plots shown in this chapter. The leading tag-jet p_{T} appears to provide good discrimination over backgrounds and a baseline cut of $> 40\,\mathrm{GeV}/c$ is implemented.

As discussed in Section 1.4.1 a characteristic feature of VBF is that the tag-jets tend to be in the forward regions of the detector and well separated in pseudo-rapidity. Therefore the tag-jets are required to be in opposite hemispheres of the detector (i.e. the product of the η of the two tag-jets ($\eta \times \eta$ (tag-jets)) must be negative (see Figure 5.12(b))). In addition, a cut on the $\Delta\eta$ (see Figure 5.12(c)) between the tag-jets is made and required to be greater than 3.0. The result of the large pseudo-rapidity separation between tag-jets is that the invariant mass of the tag-jets, shown in Figure 5.12(d) tends to be large compared to all backgrounds. The invariant mass of the two tag-jets is required to be greater than $400\,\mathrm{GeV}/c^2$ in the low mass selection. Due to the large background suppression provided by the larger possible $E_{\mathrm{T}}^{\mathrm{miss}}$ cut, no cut on the mass of the tag-jets is made in the high mass selection.

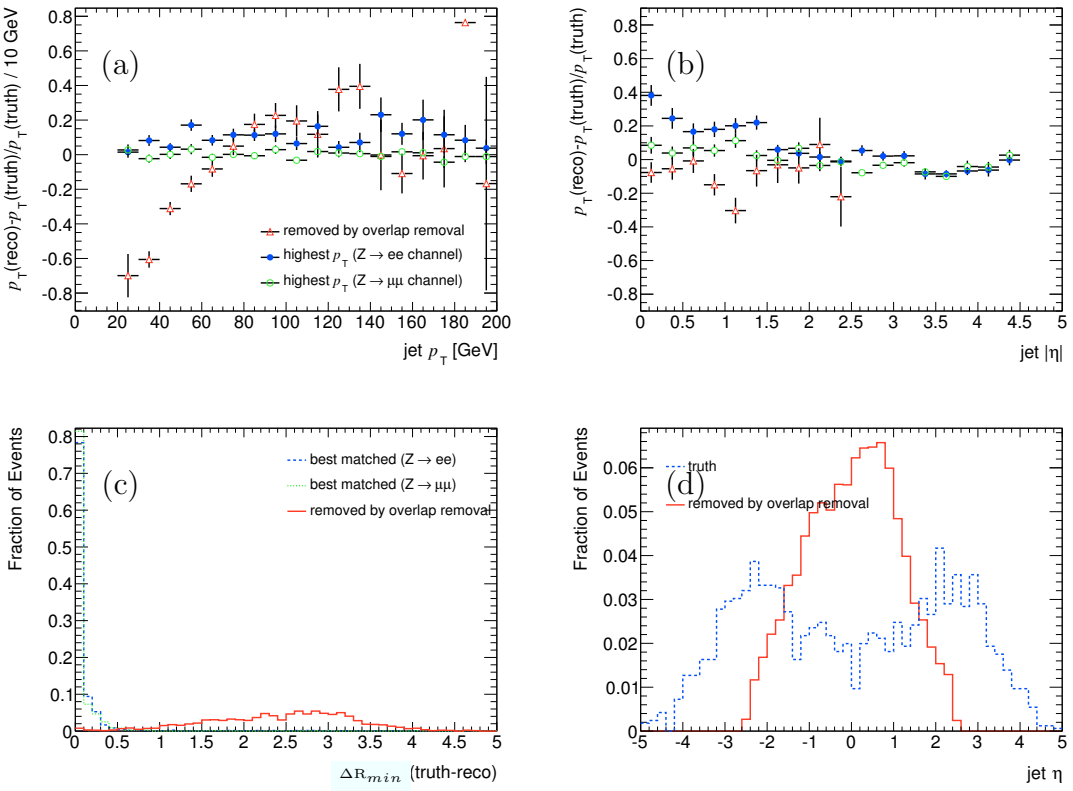


Figure 5.11: p_{T}(truth)-p_{T}(reco)/p_{T}(truth) plotted as a function of truth VBF quark (a) p_{T} and (b) $|\eta|$ for the reconstructed jets best matched to the truth VBF quarks and the jets removed by the overlap removal procedure. (c) shows the ΔR between truth VBF quarks and best matched/overlap removed reconstructed jets and (d) the η of truth VBF quarks and overlap removed tag-jets (the asymmetry of this distribution for the overlap removed tag-jets is caused by the OTX cut in the electron selection). Signal corresponds to the $\mathrm{m_H} = 200\,\mathrm{GeV}/c^2$ Higgs sample.

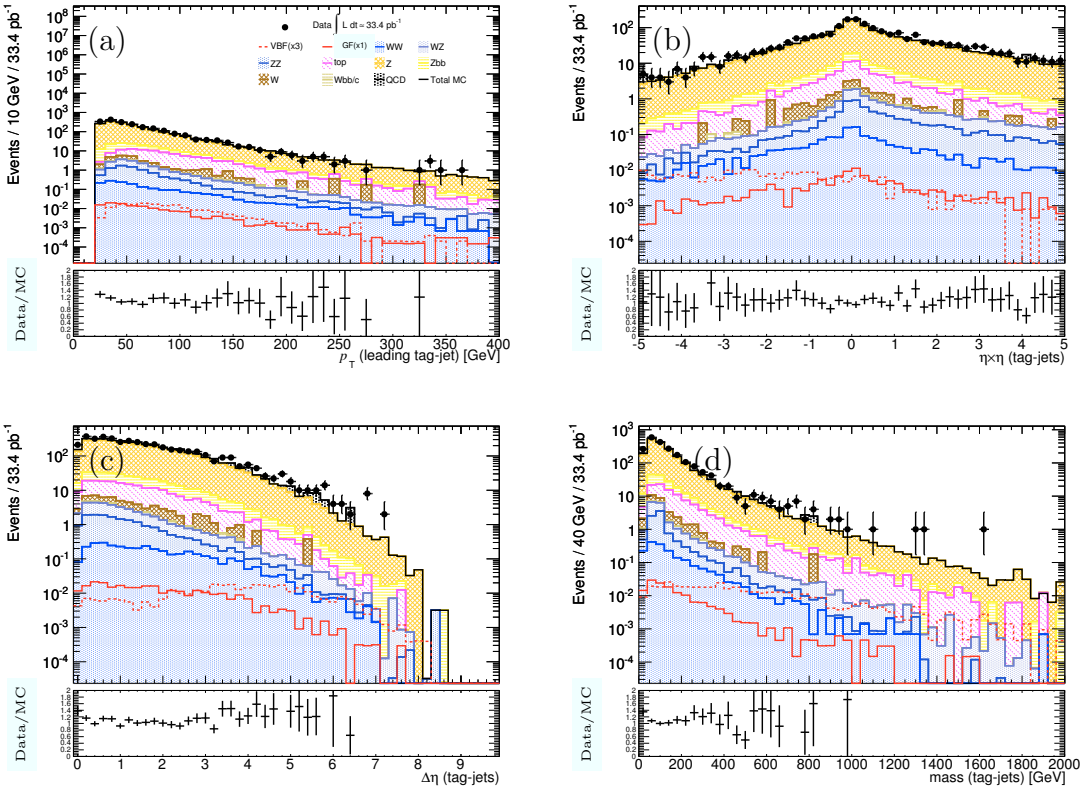


Figure 5.12: Distributions related to the kinematics of the tag-jets, (a) p_{T} of leading tag-jet, (b) $\eta \times \eta$ of tag-jets, (c) $\Delta\eta$ between tag-jets and (d) mass of tag-jets. Each background contribution is shown stacked. VBF and GF components for the $\mathrm{m_H} = 200\,\mathrm{GeV}/c^2$ are included.

Central Jet Veto The absence of colour exchange between the partons within the VBF process means that jet activity in the signal in the central region is suppressed. Therefore it is beneficial to reject(veto) events with additional jets (referred to here as non tag-jets). This cut is commonly called a central jet veto (CJV). This is motivated in Figure 5.13(a) which shows the number of additional jets with a $p_{\mathrm{T}} > 20\,\mathrm{GeV}/c$ within $|\eta| < 3.2$. For completeness the distributions of the non tag-jets p_{T} and η are shown in Figures 5.13(b) and 5.13(c). In the low mass selection a CJV cut is adopted, removing events with non tag-jets with $p_{\mathrm{T}} > 20\,\mathrm{GeV}/c$ and $|\eta| < 3.2$.

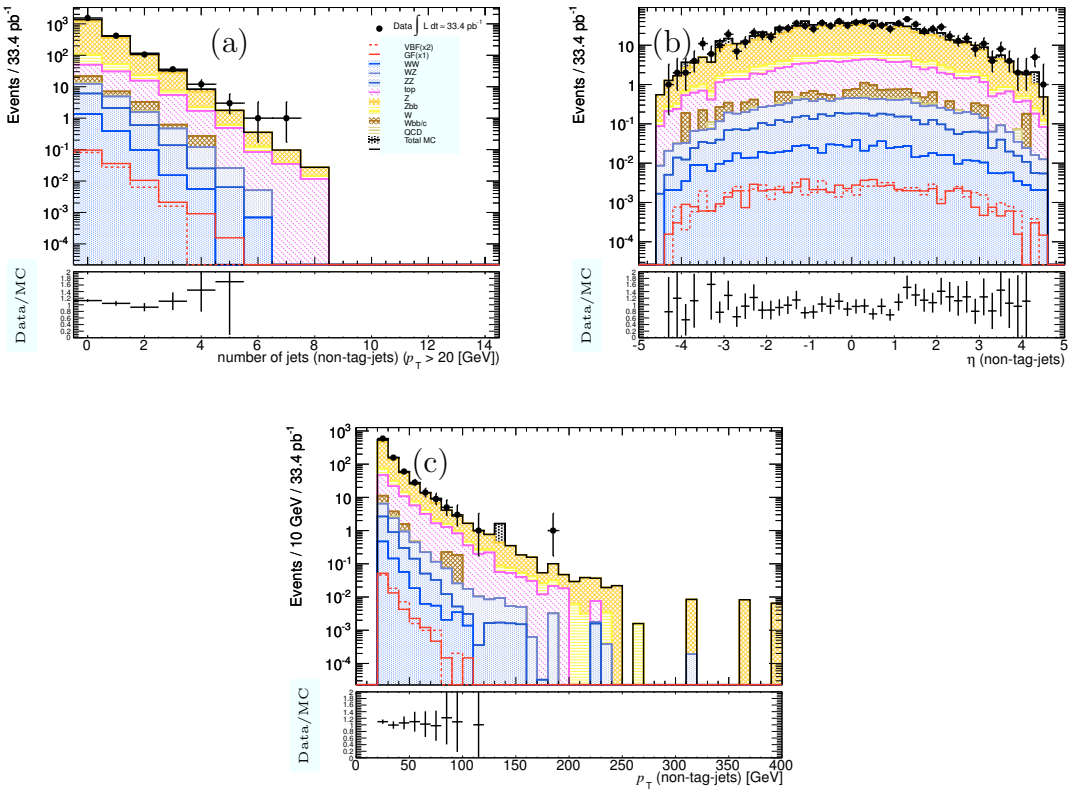


Figure 5.13: Distributions of non tag-jets (a) multiplicity, (b) η and(c) p_{T}. Each background contribution is shown stacked. VBF and GF components for the $\mathrm{m_H} = 200\,\mathrm{GeV}/c^2$ are included.

After requiring at least two jets passing the jet selection, the GF component of the signal is reduced by 70 % for the $\mathrm{m_H} = 200\,\mathrm{GeV}/c^2$ signal when compared to the yield at the end of the selection relating to the Higgs decay products while the VBF component is reduced by 25 %. The backgrounds suppressed the most correspond to those without hard jets, namely the di-boson backgrounds which reduce by as much as 90 %. However, the main background Z+jets is further reduced by over 50 % in the low mass selection. This reduction is less pronounced for the high mass selection. It is feasible this is because of the larger $E_{\mathrm{T}}^{\mathrm{miss}}$ cut only allowing Z events with large

jet activity to survive. W+jets background which was the 2^{nd} largest background at the end of the selection relating to the Higgs decay products in both the low and high mass selections is now reduced by over 80 % in each case by applying this cut, becoming a minor background. Top background from $t\bar{t}$ does not experience a major drop in efficiency and together with Z+jets forms over 80 % of the total background in both the low and high mass selections. At this stage in the selection data and Monte Carlo are not in agreement, data having approximately double the number of events than found in the Monte Carlo. This is reflected in Figure 5.14(b) which compares the multiplicity of jets satisfying the jet selection for the nominal Monte Carlo and the pile-up reweighted Monte Carlo. It shows the pile-up Monte Carlo predicts a larger contribution than the nominal Monte Carlo for jet multiplicities between one and three, verifying that the discrepancy observed between data and Monte Carlo after this cut is due to pile-up.

The leading tag-jet p_{T} cut in both the low and high mass selections is most effective at reducing W+jets background, which is reduced by approximately 50 %. All other backgrounds are affected similarly and reduce by around 10 % or less. Z+jets and top remain the main backgrounds constituting around 50 and 25 % of the total background respectively in both the low and high mass selections.

Despite reducing signal efficiency by around 1/3 in the low mass selection (there is negligible loss in the high mass selection), the total background is reduced by over twice this amount by the requirement that the tag-jets lie in opposite hemispheres. This gives an increase in significance of around 20 % in both the low and high mass selections. It is interesting to note that data and nominal Monte Carlo now agree again within statistical error. It is possible this occurs because in the current selections, the events selected by the leading tag-jet and opposite hemispheres requirement tend to select events where there is agreement between data and Monte Carlo in terms of the jet multiplicities, corresponding to events with four or more jets.

The $\Delta\eta$ cut provides a further large reduction in total background, by as much as 80 % in both selections. This is largely due to the strong suppression of what still remains the largest background, Z+jets. The signal is largely unaffected in each selection, resulting in an increase in signal significance by over 50 % in each case. At this stage in the selection, just one data event remains in the low mass selection while no events survive the high mass selection. The discrepancy between data and Monte Carlo is still largely due to pile-up effects. This is verified in Figures 5.14 and 5.15 which compare the tag-jet variable distributions for non pile-up and pile-up reweighted Monte Carlo. The distributions in Figures 5.14(b) and 5.14(h) corresponding to the variables $\mathrm{N_{jets}}$ and $\Delta\eta$ are shown not to agree, with the non pile-up Monte Carlo underestimating the pile-up Monte Carlo. However, for the

other variables considered pile-up and non pile-up Monte Carlo are observed to agree within error.

As expected the CJV has little effect other than on those backgrounds expected to have a non-negligible amount of jet activity, namely $t\bar{t}$ and Z+jets background. The largest decrease is seen for $t\bar{t}$ at approximately 50 %. A negligible increase in signal significance is observed.

The cut on the mass of the tag-jets removes the remaining W statistics and leads to a reduction of the Z+jets and $t\bar{t}$ backgrounds by 60 % or more in each selection contributing to a increase in signal significance by around 40 %.

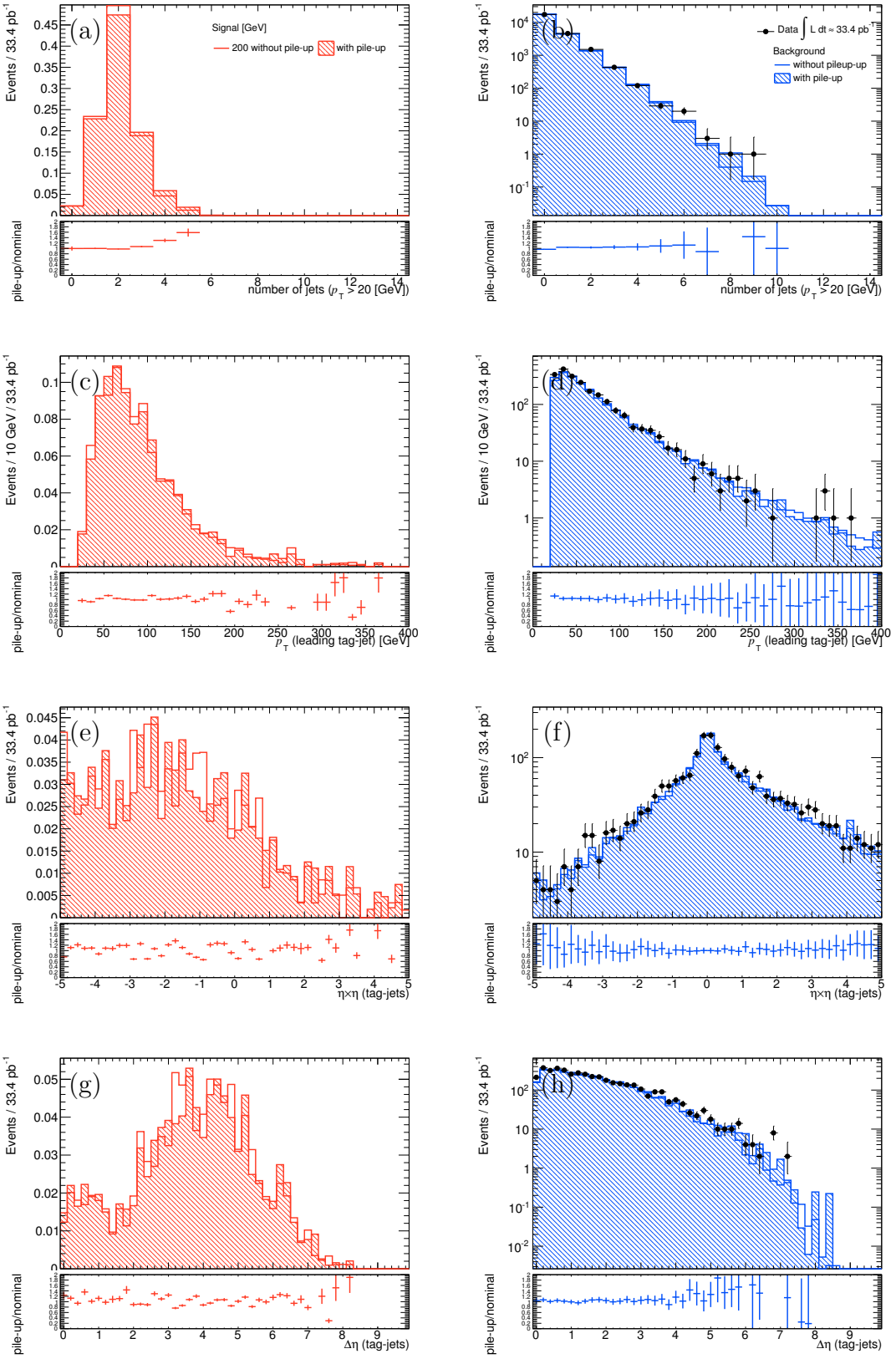


Figure 5.14: Comparison of non pile-up and pile-up reweighted Monte Carlo for $m_H = 200\,\mathrm{GeV}/c^2$ signal and total background for (a,b) jet multiplcity, (c,d) leading tag-jet p_T, (e,f) $\eta \times \eta$ and (g,h) $\Delta\eta$ between tag-jets.

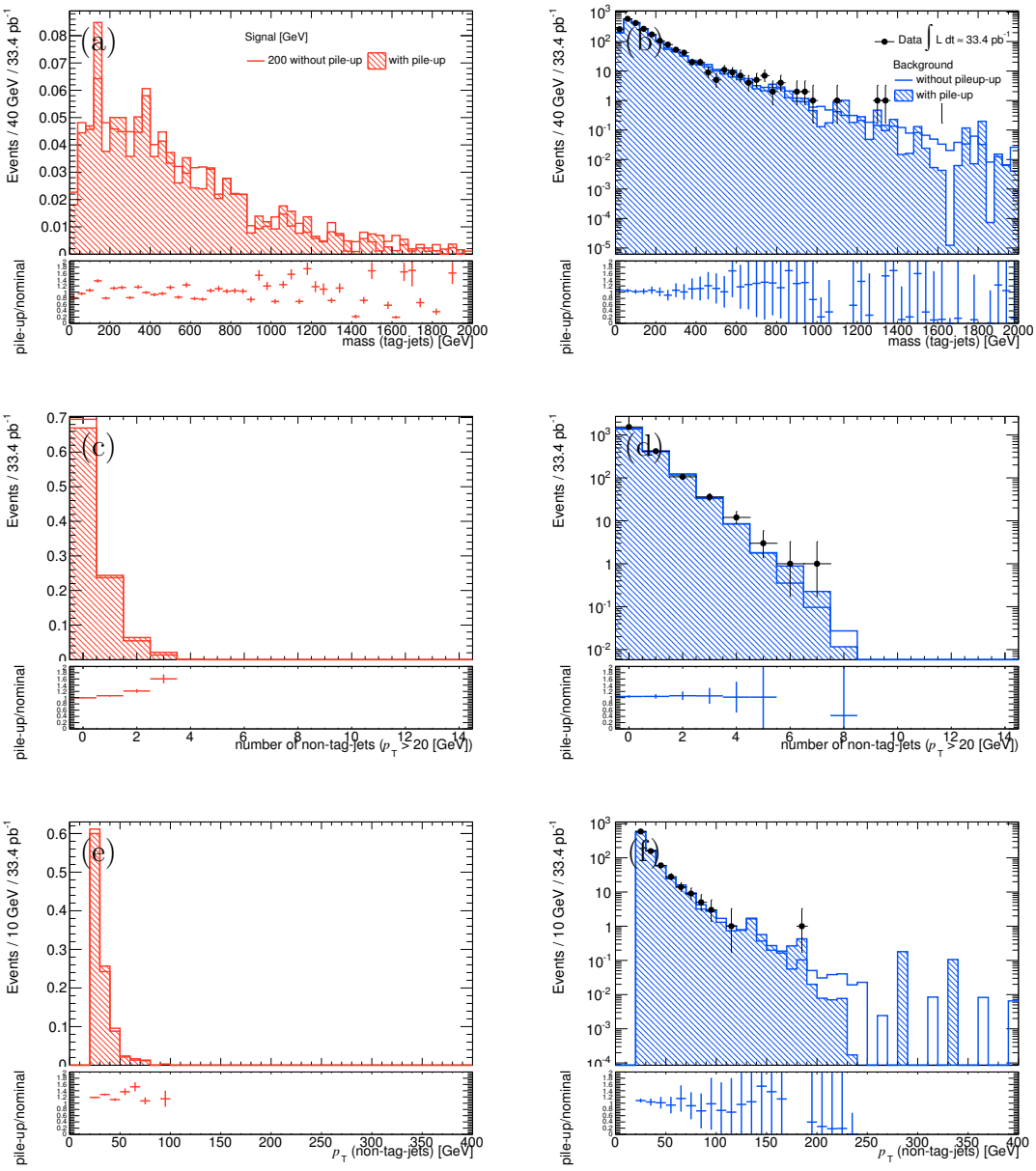


Figure 5.15: Comparison of non pile-up and pile-up reweighted Monte Carlo for $m_H = 200\,\mathrm{GeV}/c^2$ signal and total background for (a,b) mass tag-jets, (c,d) non tag-jet multiplicity and (e,f) non tag-jet p_T.

Results of baseline selection

In this section a baseline analysis was presented for the cut-based selection of the VBF H $\rightarrow$ ZZ $\rightarrow$ $ll\nu\nu$ signal, relating to the VBF remnants. The selection is summarised in the following. Where there is a difference between low and high mass selections this is indicated with the low mass cut value defined and next to it the high mass cut value in brackets.

- $N_{jets} \geq 2$: At least two jets passing the standard selection (see Section 4.4.1).

- leading tag-jet p_{T}: $> 40\,\mathrm{GeV}/c$
- $\eta \times \eta$: Product of tag-jets absolute η must be less than 0, i.e. tag-jets must be in opposite hemispheres.
- $\Delta\eta$: > 3.0(no cut).
- CJV: Reject events with non tag-jets with $p_{\mathrm{T}} > 20\,\mathrm{GeV}/c$ and $|\eta| < 2.5$ (no cut).
- mass tag-jets: > 400(no cut) GeV/c^2.

The number of simulated events and events normalised to the luminosity analysed in the data passing these selection cuts for the low and high mass selections are presented in Tables 5.7-5.8. As is shown in these tables, a number of samples have zero Monte Carlo events after the selection detailed, including QCD, W+jets, single top and Z$\rightarrow \tau\tau$+jets. Other samples have zero Monte Carlo events surviving for particular physical processes. The rightmost columns in the Tables shown in Table 5.8 shows the effect on the final expected background rate when the zero surviving Monte Carlo events in these samples is replaced by one. Although this method will overestimate the background, as in all cases the final efficiency of each sample is less than the inverse of the total number of initial Monte Carlo events, it does provide a crude means by which background contributions which might otherwise be neglected can be approximated. The total in this column is calculated using samples which did not completely run out of Monte Carlo events when the full selection was applied, and as such can be compared to that at the end of the full selection. It is seen that the increase in background rate observed is within statistical error. However, for samples where no Monte Carlo events remain after the selection, potentially large increases in background rate are predicted using this method, particularly for QCD and W. Therefore an attempt is made to try and estimate their effect in the following section.

Baseline Low Mass Selection (Z → $\mu\mu/ee$ channels combined)							
Sample	Initial	$N_{jets} \geq 2$	leading tag-jet p_T	$\eta \times \eta$	$\Delta\eta$	CJV	mass tag-jets
VBF	614	480	450	369	301	259	206
GF	1679	479	389	185	89	81	40
$t\bar{t}$	811	692	625	273	78	43	14
WW	2900	418	332	166	61	53	20
ZZ	13483	1976	1406	604	191	175	59
WZ	9320	2968	2306	961	293	230	93
Z+jets	2181	1111	1056	466	74	51	21
W+jets	25	7	4	3	2	2	0
Wbb+jets	9	3	3	1	1	0	0
Zbb+jets	623	452	396	166	30	19	6
QCD	3	1	1	0	0	0	0
single top	10	4	4	2	0	0	0
Z→ $\tau\tau$+jets	36	10	9	4	1	1	0
Baseline High Mass Selection (Z → $\mu\mu/ee$ channels combined)							
Sample	Initial	$N_{jets} \geq 2$	leading tag-jet p_T	$\eta \times \eta$	$\Delta\eta$	CJV	mass tag-jets
VBF	894	679	635	511	440	440	440
GF	3133	828	635	308	134	134	134
$t\bar{t}$	258	215	189	90	20	20	20
WW	945	116	90	39	16	16	16
ZZ	6302	877	634	267	78	78	78
WZ	3530	1168	923	385	101	101	101
Z+jets	444	331	327	138	18	18	18
W+jets	6	1	0	0	0	0	0
Wbb+jets	0	0	0	0	0	0	0
Zbb+jets	151	124	119	52	9	9	9
QCD	0	0	0	0	0	0	0
single top	1	1	1	0	0	0	0
Z→ $\tau\tau$+jets	2	1	1	0	0	0	0

Table 5.7: Number of simulated events passing baseline selection relating to VBF remnants for the low and high mass selections (Section 5.2). Included are combined statistics for the Z → $\mu\mu/ee$ channels. The signal corresponds to the m_H = 200(300) GeV/c^2 sample for the low(high) mass selection.

Baseline Low Mass Selection ($Z \to \mu\mu/ee$ channels combined)								
Sample	Initial	$N_{jets} \geq 2$	leading tag-jet p_T	$\eta \times \eta$	$\Delta\eta$	CJV	mass tag-jets	1Event(Total)
VBF	0.04±0.002 (0.19)	0.03±0.001 (0.15)	0.03±0.001 (0.14)	0.02±0.001 (0.11)	0.02±0.001 (9.2×10^{-2})	0.02±0.001 (7.9×10^{-2})	0.01±0.001 (6.3×10^{-2})	0.01±0.001
GF	0.26±0.0064 (0.14)	0.07±0.003 (4.1×10^{-2})	0.06±0.003 (3.4×10^{-2})	0.03±0.002 (1.6×10^{-2})	0.01±0.001 (7.7×10^{-3})	0.01±0.001 (7.0×10^{-3})	0.006±0.001 (3.5×10^{-3})	0.006±0.001
$t\bar{t}$	4.97±0.17 (1.6×10^{-3})	4.24±0.16 (1.4×10^{-3})	3.83±0.15 (1.3×10^{-3})	1.67±0.10 (5.5×10^{-4})	0.48±0.05 (1.6×10^{-4})	0.26±0.04 (8.5×10^{-5})	0.09±0.02 (2.9×10^{-5})	0.09±0.02
WW	2.07±0.04 (1.3×10^{-2})	0.30±0.02 (1.8×10^{-3})	0.24±0.01 (1.5×10^{-3})	0.12±0.01 (7.3×10^{-4})	0.04±0.007 (2.6×10^{-4})	0.04±0.006 (2.3×10^{-4})	0.01±0.005 (9.1×10^{-5})	0.02±0.006
ZZ	1.42±0.02 (3.6×10^{-2})	0.23±0.0094 (5.8×10^{-3})	0.17±0.0088 (4.4×10^{-3})	0.07±0.007 (1.8×10^{-3})	0.02±0.004 (5.0×10^{-4})	0.02±0.004 (4.6×10^{-4})	0.006±0.008 (1.5×10^{-4})	0.008±0.004
WZ	1.63±0.03 (1.2×10^{-2})	0.56±0.02 (4.2×10^{-3})	0.44±0.02 (3.3×10^{-3})	0.18±0.0085 (1.3×10^{-3})	0.05±0.006 (4.0×10^{-4})	0.04±0.005 (3.1×10^{-4})	0.02±0.002 (1.2×10^{-4})	0.02±0.01
Z+jets	17.51±0.38 (2.5×10^{-4})	8.80±0.27 (1.3×10^{-4})	8.36±0.26 (1.2×10^{-4})	3.66±0.17 (5.3×10^{-5})	0.58±0.08 (8.4×10^{-6})	0.41±0.07 (6.0×10^{-6})	0.17±0.06 (2.4×10^{-6})	0.21±0.07
W+jets	6.12±1.96 (8.9×10^{-6})	1.09±0.88 (1.6×10^{-6})	0.62±0.67 (9.1×10^{-7})	0.46±0.56 (6.7×10^{-7})	0.31±0.52 (4.5×10^{-7})	0.31±0.52 (4.5×10^{-7})	-	3.83±2.24
Wbb+jets	0.35±0.21 (1.3×10^{-5})	0.11±0.13 (4.1×10^{-6})	0.11±0.13 (4.1×10^{-6})	0.04±0.09 (1.3×10^{-6})	0.04±0.09 (1.3×10^{-6})	-	-	0.35±0.27
Zbb+jets	1.50±0.06 (1.8×10^{-3})	1.00±0.05 (1.2×10^{-3})	0.87±0.05 (1.1×10^{-3})	0.37±0.03 (4.5×10^{-4})	0.07±0.02 (8.0×10^{-5})	0.04±0.02 (4.4×10^{-5})	0.01±0.01 (1.4×10^{-5})	0.02±0.02
QCD	2.12±2.88 (6.0×10^{-12})	0.002±0.004 (4.5×10^{-15})	0.002±0.004 (4.5×10^{-15})	-	-	-	-	505119±1078932
single top	0.24±0.12 (5.4×10^{-4})	0.11±0.09 (2.5×10^{-4})	0.11±0.09 (2.5×10^{-4})	0.06±0.07 (1.3×10^{-4})	-	-	-	0.07±0.09
$Z\to\tau\tau$+jets	0.29±0.06 (8.2×10^{-6})	0.07±0.04 (2.1×10^{-6})	0.06±0.04 (1.8×10^{-6})	0.03±0.03 (9.1×10^{-7})	0.008±0.02 (2.2×10^{-7})	0.008±0.02 (2.2×10^{-7})	-	0.06±0.05
Total	38.22±3.52 (1.08×10^{-10})	16.51±0.95 (4.67×10^{-11})	14.82±0.76 (4.20×10^{-11})	6.66±0.61	1.59±0.54 (4.51×10^{-12})	1.12±0.53 (3.18×10^{-12})	0.30±0.06 (8.59×10^{-13})	0.37±0.08
Data	41	25	23	7	1	1	1	
$s/\sqrt{b}$	0.007±0.0003	0.008±0.0004	0.008±0.0004	0.01±0.0005	0.016±0.002	0.017±0.004	0.025±0.005	0.023±0.01

Baseline High Mass Selection ($Z \to \mu\mu/ee$ channels combined)								
Sample	Initial	$N_{jets} \geq 2$	leading tag-jet p_T	$\eta \times \eta$	$\Delta\eta$	CJV	mass tag-jets	1Event(Total)
VBF	0.03±0.001 (0.26)	0.03±0.001 (0.20)	0.02±0.0009 (0.19)	0.02±0.0008 (0.15)	0.02±0.0008 (0.13)	0.02±0.0008 (0.13)	0.02±0.0008 (0.13)	0.02±0.0008
GF	0.27±0.0049 (0.27)	0.07±0.003 (7.2×10^{-2})	0.06±0.002 (5.5×10^{-2})	0.03±0.002 (2.7×10^{-2})	0.01±0.001 (1.2×10^{-2})	0.01±0.001 (1.2×10^{-2})	0.01±0.001 (1.2×10^{-2})	0.01±0.001
$t\bar{t}$	1.58±0.10 (5.2×10^{-4})	1.32±0.09 (4.3×10^{-4})	1.16±0.08 (3.8×10^{-4})	0.55±0.06 (1.8×10^{-4})	0.13±0.03 (4.1×10^{-5})	0.13±0.03 (4.1×10^{-5})	0.13±0.03 (4.1×10^{-5})	0.13±0.03
WW	0.67±0.02 (4.1×10^{-3})	0.08±0.009 (5.1×10^{-4})	0.06±0.008 (3.9×10^{-4})	0.03±0.006 (1.7×10^{-4})	0.01±0.005 (7.2×10^{-5})	0.01±0.005 (7.2×10^{-5})	0.01±0.005 (7.2×10^{-5})	0.01±0.006
ZZ	0.67±0.01 (1.7×10^{-2})	0.10±0.007 (2.6×10^{-3})	0.08±0.007 (1.9×10^{-3})	0.03±0.005 (8.3×10^{-4})	0.009±0.004 (2.3×10^{-4})	0.009±0.004 (2.3×10^{-4})	0.009±0.004 (2.3×10^{-4})	0.01±0.004
WZ	0.62±0.01 (4.6×10^{-3})	0.22±0.0088 (1.6×10^{-3})	0.17±0.0081 (1.3×10^{-3})	0.07±0.007 (5.4×10^{-4})	0.02±0.004 (1.5×10^{-4})	0.02±0.004 (1.5×10^{-4})	0.02±0.004 (1.5×10^{-4})	0.03±0.01
Z+jets	3.55±0.17 (5.1×10^{-5})	2.63±0.15 (3.8×10^{-5})	2.59±0.15 (3.8×10^{-5})	1.09±0.10 (1.6×10^{-5})	0.14±0.06 (2.0×10^{-6})	0.14±0.06 (2.0×10^{-6})	0.14±0.06 (2.0×10^{-6})	0.19±0.07
W+jets	1.63±1.26 (2.4×10^{-6})	0.15±0.37 (2.3×10^{-7})	-	-	-	-	-	3.83±2.24
Wbb+jets	-	-	-	-	-	-	-	0.35±0.27
Zbb+jets	0.36±0.03 (4.3×10^{-4})	0.28±0.03 (3.3×10^{-4})	0.26±0.03 (3.2×10^{-4})	0.12±0.02 (1.4×10^{-4})	0.02±0.01 (2.2×10^{-5})	0.02±0.01 (2.2×10^{-5})	0.02±0.01 (2.2×10^{-5})	0.03±0.02
QCD	-	-	-	-	-	-	-	505119±1078932
single top	0.03±0.07 (6.4×10^{-5})	0.03±0.07 (6.4×10^{-5})	0.03±0.07 (6.4×10^{-5})	-	-	-	-	0.07±0.09
$Z\to\tau\tau$+jets	0.01±0.02 (2.9×10^{-7})	0.002±0.005 (5.5×10^{-8})	0.002±0.005 (5.5×10^{-8})	-	-	-	-	0.06±0.05
Total	9.11±1.28 (2.58×10^{-11})	4.81±0.41 (1.36×10^{-11})	4.36±0.18 (1.23×10^{-11})	1.89±0.12	0.33±0.06 (9.23×10^{-13})	0.33±0.06 (9.23×10^{-13})	0.33±0.06 (9.23×10^{-13})	0.39±0.08
Data	9	7	7	2	0	0	0	
$s/\sqrt{b}$	0.011±0.0004	0.011±0.0005	0.011±0.0005	0.014±0.0007	0.028±0.005	0.028±0.005	0.028±0.005	0.026±0.01

Table 5.8: Expected number of simulated events passing baseline selection based on VBF remnants using low/ high mass cuts given an integrated luminosity of $33.4\,\mathrm{pb}^{-1}$. Included are the combined statistics for the $Z \to \mu\mu/ee$ channels. Efficiencies are given in brackets. The signal corresponds to the $m_H = 200(300)\,\mathrm{GeV}/c^2$ sample for the low(high) mass selection respectively.

5.4 Estimation of QCD, W, Z→ ττ+jets and single top contributions

It was shown in Section 5.3 that the contributions from the QCD, W, Z → ττ+jets and single top backgrounds could not be estimated very reliably due to lack of Monte Carlo statistics. Here we use an alternative approach to estimate these background contributions.

The approach used to estimate their contribution, called factorization, relies on the property that if a set of cuts A are independent to those in set B, then the efficiency after applying cuts A and B together is the same as multiplying the efficiency found by applying cuts A independently of set B by the efficiency of applying cuts B independently of cuts A as shown in Eqn. 5.2.

$$\epsilon_{AB} = \epsilon_A \times \epsilon_B \tag{5.2}$$

In general, this holds for any number of sets of cuts provided each set is independent of the other. In the following three different sets of cuts are identified as independent in the analysis, those relating to the lepton cuts, the $E_{\mathrm{T}}^{\mathrm{miss}}$ cut and the tag-jet cuts

- Lepton cuts: relaxed lepton selection and di-lepton mass window.
- $E_{\mathrm{T}}^{\mathrm{miss}}$ cut: $E_{\mathrm{T}}^{\mathrm{miss}} > 40$ GeV.
- tag-jet cuts: $\mathrm{N_{jets}} \geq 2$, each with p_{T} >20 GeV/c $|\eta| < 4.5$. Highest p_{T} jets $\eta \times \eta$ and $\Delta\eta$ as in baseline selection.

The relaxed lepton selection requires either two medium electrons or two combined or segment tagged muons (one of which must be combined). In the electron channel the electron cluster must have $E_{\mathrm{T}} > 20$ GeV and the electron track $|\eta| < 2.5$. In the muon channel, the muon is required to have $p_{\mathrm{T}} > 20$ GeV/c and $|\eta| < 2.5$. This more relaxed set of cuts is used to maximise the acceptance of the samples considered (particularly QCD) and means that the estimated contributions are expected to be less than the values determined.

The efficiencies of applying each individual set of cuts on each sub-sample making up the QCD, W+jets, Z → ττ+jets and single top backgrounds are computed. All the available statistics have been used in each case. The contribution to the total background of each sub-sample is then estimated by finding the product of efficiencies from applying each set of cuts and using this to find the expected number of events for each sub-sample.

Sample		No Cuts	Lepton Cuts	$E_{\mathrm{T}}^{\mathrm{miss}}$ cut	tag-jet cuts	Combined
J0	MC	1.39918×10^{6}	$0\to 1$	1	5	
	ϵ	1	7.15×10^{-7}	7.15×10^{-7}	3.57×10^{-6}	1.83×10^{-18}
	n	3.29×10^{11}	$< 2.35\times10^{5}$	2.35×10^{5}	1.18×10^{6}	6.01×10^{-7}
J1	MC	1.40×10^{6}	$0\to 1$	3	486	
	ϵ	1	7.17×10^{-7}	2.15×10^{-6}	3.48×10^{-4}	5.37×10^{-16}
	n	2.26e+10	$< 1.62\times10^{4}$	4.87×10^{4}	7.89×10^{6}	1.22×10^{-5}
J2	MC	1.40×10^{6}	$0\to 1$	118	4430	
	ϵ	1	7.15×10^{-7}	8.44×10^{-5}	3.17×10^{-3}	1.91×10^{-13}
	n	1.37e+09	$< 9.79\times10^{2}$	1.16×10^{5}	4.34×10^{6}	2.62×10^{-4}
J3	MC	1.40×10^{6}	6	4628	12297	
	ϵ	1	4.29×10^{-6}	3.31×10^{-3}	8.80×10^{-3}	1.25×10^{-10}
	n	7.33×10^{7}	3.15×10^{2}	2.43×10^{5}	6.45×10^{5}	9.17×10^{-3}
J4	MC	1.40×10^{6}	29	47525	7555	
	ϵ	1	2.08×10^{-5}	3.40×10^{-2}	5.41×10^{-3}	3.82×10^{-9}
	n	2.93×10^{6}	6.08×10^{1}	9.96×10^{4}	1.58×10^{4}	1.12×10^{-2}
J5	MC	1.39×10^{6}	92	203795	1316	
	ϵ	1	6.61×10^{-5}	1.46×10^{-1}	9.46×10^{-4}	9.16×10^{-9}
	n	7.85×10^{4}	5.19	1.15×10^{4}	7.42×10^{1}	7.19×10^{-4}
J6	MC	1.35×10^{6}	210	492457	11	
	ϵ	1	1.56×10^{-4}	3.65×10^{-1}	8.16×10^{-6}	4.65×10^{-10}
	n	1.12×10^{3}	1.75×10^{-1}	4.10×10^{2}	9.16×10^{-3}	5.22×10^{-7}
J7	MC	1.13×10^{6}	545	668556	2	
	ϵ	1	4.84×10^{-4}	5.94×10^{-1}	1.78×10^{-6}	5.11×10^{-10}
	n	4.59	2.22×10^{-3}	2.73	8.16×10^{-6}	2.35×10^{-9}
J8	MC	1.38×10^{6}	734	843653	1	
	ϵ	1	5.32×10^{-4}	6.11×10^{-1}	7.25×10^{-7}	2.35×10^{-10}
	n	2.08×10^{-4}	1.11×10^{-7}	1.27×10^{-4}	1.51×10^{-10}	4.90×10^{-14}
bb						
$\to$ ee	MC	4.45×10^{6}	88	21860	4355	
	ϵ	1	1.98×10^{-5}	4.91×10^{-3}	9.79×10^{-4}	9.52×10^{-11}
	n	2.47×10^{6}	4.88×10^{1}	1.21×10^{4}	2.42×10^{3}	2.35×10^{-4}
$\to \mu\mu$	MC	4.44×10^{6}	453	155025	3352	
	ϵ	1	1.02×10^{-4}	3.49×10^{-2}	7.54×10^{-4}	2.68×10^{-9}
	n	2.47×10^{6}	2.52×10^{2}	8.61×10^{4}	1.86×10^{3}	6.62×10^{-3}
cc						
$\to$ ee	MC	1.49×10^{6}	16	7172	3184	
	ϵ	1	1.07×10^{-5}	4.80×10^{-3}	2.13×10^{-3}	1.09×10^{-10}
	n	9.49×10^{5}	1.02×10^{1}	4.55×10^{3}	2.02×10^{3}	1.04×10^{-4}
$\to \mu\mu$	MC	1.50×10^{6}	90	52748	2083	
	ϵ	1	6.00×10^{-5}	3.52×10^{-2}	1.39×10^{-3}	2.93×10^{-9}
	n	9.49×10^{5}	5.69×10^{1}	3.34×10^{4}	1.32×10^{3}	2.78×10^{-3}
Total	n	3.53×10^{11}	2.53×10^{5}	8.90×10^{5}	1.41×10^{7}	3.11×10^{-2}

Table 5.9: Results from using the cut factorisation method to estimate the contribution of QCD. The rows represent each sub-sample considered, with the bottom row the total QCD contribution. Moving from left to right, the columns correspond to no cuts applied, lepton cuts, $E_{\mathrm{T}}^{\mathrm{miss}}$ cut, tag-jet cuts and the combined effect of each using the factorisation method. Number of simulated events is shown by MC, corresponding number of events normalised to luminosity used in this analysis by n and the corresponding efficiency, by ϵ.

Table 5.9 shows the result of applying this procedure to each sub-sample associated with the QCD background. The *No cuts* column shows the initial number of simulated events (MC), the corresponding number of events normalised to luminosity used in this analysis (n) (luminosity in data $\approx 33.4\,\mathrm{pb}^{-1}$) and the corresponding efficiency (ϵ). The effect of applying each individual set of cuts (lepton cuts, $E_{\mathrm{T}}^{\mathrm{miss}}$ cut and tag-jet cuts) is then shown in the subsequent columns in terms MC,n and ϵ. Where no Monte Carlo events survive the cuts, the contribution is calculated assuming one simulated event survives. The *Combined* column represents the expected contribution from each sub-sample using the factorisation method. The bottom row of the table shows the total contribution for each case. The results from estimating the W+jets, Z $\rightarrow \tau\tau$+jets and single top contributions are shown in Tables 5.10-5.12. Overall the contribution from QCD is expected to be less than 3.1×10^{-2} events and that from W less than 3.5×10^{-2} events. The contribution from Z $\rightarrow \tau\tau$+jets and single top is much less at 7.6×10^{-4} and 2.0×10^{-3} respectively, making the total contribution of these samples within the statistical error on the total background.

Sample		No Cuts	Lepton Cuts	$E_{\mathrm{T}}^{\mathrm{miss}}$ cut	tag-jet cuts	Combined
W → eν						
	MC	1.38193×10^{6}	0→1	88979	610	
+ 0 jets	ϵ	1	7.24×10^{-7}	6.44×10^{-2}	4.41×10^{-4}	2.06×10^{-11}
	n	1.55×10^{5}	$< 1.12\times10^{-1}$	9.98×10^{3}	6.84×10^{1}	3.19×10^{-6}
	MC	2.58×10^{5}	77	65282	108	
+ 1 jets	ϵ	1	2.98×10^{-4}	2.53×10^{-1}	4.18×10^{-4}	3.15×10^{-8}
	n	5.27×10^{4}	1.57×10^{1}	1.33×10^{4}	2.20×10^{1}	1.66×10^{-3}
	MC	1.89×10^{5}	108	60978	483	
+ 2 jets	ϵ	1	5.72×10^{-4}	3.23×10^{-1}	2.56×10^{-3}	4.72×10^{-7}
	n	1.54×10^{4}	8.79	4.96×10^{3}	3.93×10^{1}	7.25×10^{-3}
	MC	5.05×10^{4}	32	19409	294	
+ 3 jets	ϵ	1	6.34×10^{-4}	3.85×10^{-1}	5.82×10^{-3}	1.42×10^{-6}
	n	4.11×10^{3}	2.61	1.58×10^{3}	2.39×10^{1}	5.84×10^{-3}
	MC	1.30×10^{4}	13	5754	109	
+ 4 jets	ϵ	1	1.00×10^{-3}	4.43×10^{-1}	8.39×10^{-3}	3.72×10^{-6}
	n	1.03×10^{3}	1.03	4.57×10^{2}	8.65	3.83×10^{-3}
	MC	3.45×10^{3}	3	1650	52	
+ 5 jets	ϵ	1	8.70×10^{-4}	4.78×10^{-1}	1.51×10^{-2}	6.27×10^{-6}
	n	2.81×10^{2}	2.45×10^{-1}	1.35×10^{2}	4.24	1.76×10^{-3}
W → μν						
	MC	1.39×10^{6}	1	689	26	
+ 0 jets	ϵ	1	7.21×10^{-7}	4.97×10^{-4}	1.88×10^{-5}	6.73×10^{-15}
	n	1.55×10^{5}	1.12×10^{-1}	7.70×10^{1}	2.91	1.04×10^{-9}
	MC	2.56×10^{5}	25	47852	55	
+ 1 jets	ϵ	1	9.77×10^{-5}	1.87×10^{-1}	2.15×10^{-4}	3.93×10^{-9}
	n	5.22×10^{4}	5.10	9.76×10^{3}	1.12×10^{1}	2.05×10^{-4}
	MC	1.88×10^{5}	51	67635	239	
+ 2 jets	ϵ	1	2.71×10^{-4}	3.60×10^{-1}	1.27×10^{-3}	1.24×10^{-7}
	n	1.53×10^{4}	4.15	5.51×10^{3}	1.95×10^{1}	1.90×10^{-3}
	MC	5.09×10^{4}	26	24602	227	
+ 3 jets	ϵ	1	5.11×10^{-4}	4.83×10^{-1}	4.46×10^{-3}	1.10×10^{-6}
	n	4.12×10^{3}	2.10	1.99×10^{3}	1.84×10^{1}	4.54×10^{-3}
	MC	1.30×10^{4}	2	7322	101	
+ 4 jets	ϵ	1	1.54×10^{-4}	5.64×10^{-1}	7.77×10^{-3}	6.75×10^{-7}
	n	1.05×10^{3}	1.61×10^{-1}	5.90×10^{2}	8.14	7.06×10^{-4}
	MC	3.50×10^{3}	1	2246	43	
+ 5 jets	ϵ	1	2.86×10^{-4}	6.42×10^{-1}	1.23×10^{-2}	2.26×10^{-6}
	n	2.85×10^{2}	8.15×10^{-2}	1.83×10^{2}	3.51	6.44×10^{-4}
W → τν						
	MC	1.37×10^{6}	0→ 1	10951	201	
+ 0 jets	ϵ	1	7.32×10^{-7}	8.02×10^{-3}	1.47×10^{-4}	8.65×10^{-13}
	n	1.55×10^{5}	$< 1.13\times10^{-1}$	1.24×10^{3}	2.28×10^{1}	1.34×10^{-7}
	MC	2.55×10^{5}	3	44681	64	
+ 1 jets	ϵ	1	1.18×10^{-5}	1.75×10^{-1}	2.51×10^{-4}	5.19×10^{-10}
	n	5.20×10^{4}	6.13×10^{-1}	9.12×10^{3}	1.31×10^{1}	2.70×10^{-5}
	MC	1.88×10^{5}	8	56675	231	
+ 2 jets	ϵ	1	4.25×10^{-5}	3.01×10^{-1}	1.23×10^{-3}	1.57×10^{-8}
	n	1.53×10^{4}	6.51×10^{-1}	4.62×10^{3}	1.88×10^{1}	2.40×10^{-4}
	MC	5.05×10^{4}	7	20242	204	
+ 3 jets	ϵ	1	1.39×10^{-4}	4.01×10^{-1}	4.04×10^{-3}	2.25×10^{-7}
	n	4.11×10^{3}	5.70×10^{-1}	1.65×10^{3}	1.66×10^{1}	9.23×10^{-4}
	MC	1.30×10^{4}	1	6260	112	
+ 4 jets	ϵ	1	7.69×10^{-5}	4.82×10^{-1}	8.62×10^{-3}	3.19×10^{-7}
	n	1.05×10^{3}	8.06×10^{-2}	5.04×10^{2}	9.03	3.35×10^{-4}
	MC	4.00×10^{3}	0→ 1	2184	51	
+ 5 jets	ϵ	1	2.50×10^{-4}	5.46×10^{-1}	1.28×10^{-2}	1.74×10^{-6}
	n	2.85×10^{2}	$< 7.13\times10^{-2}$	1.56×10^{2}	3.64	4.97×10^{-4}
Wbb						
	MC	6.50×10^{3}	8	1873	4	
+ 0 jets	ϵ	1	1.23×10^{-3}	2.88×10^{-1}	6.15×10^{-4}	2.18×10^{-7}
	n	1.30×10^{2}	1.61×10^{-1}	3.76×10^{1}	8.03×10^{-2}	2.85×10^{-5}
	MC	5.50×10^{3}	6	2257	3	
+ 1 jets	ϵ	1	1.09×10^{-3}	4.10×10^{-1}	5.45×10^{-4}	2.44×10^{-7}
	n	1.06×10^{2}	1.16×10^{-1}	4.35×10^{1}	5.78×10^{-2}	2.59×10^{-5}
	MC	3.00×10^{3}	7	1436	11	
+ 2 jets	ϵ	1	2.34×10^{-3}	4.79×10^{-1}	3.67×10^{-3}	4.11×10^{-6}
	n	5.70×10^{1}	1.33×10^{-1}	2.73×10^{1}	2.09×10^{-1}	2.34×10^{-4}
	MC	1.50×10^{3}	6	796	11	
+ 3 jets	ϵ	1	4.00×10^{-3}	5.31×10^{-1}	7.33×10^{-3}	1.56×10^{-5}
	n	2.44×10^{1}	9.78×10^{-2}	1.30×10^{1}	1.79×10^{-1}	3.81×10^{-4}
Wc						
	MC	8.63×10^{5}	156	109179	167	
+ 0 jets	ϵ	1	1.81×10^{-4}	1.27×10^{-1}	1.94×10^{-4}	4.43×10^{-9}
	n	1.76×10^{4}	3.18	2.22×10^{3}	3.40	7.79×10^{-5}
	MC	3.20×10^{5}	130	77873	595	
+ 1 jets	ϵ	1	4.06×10^{-4}	2.43×10^{-1}	1.86×10^{-3}	1.83×10^{-7}
	n	6.52×10^{3}	2.65	1.59×10^{3}	1.21×10^{1}	1.19×10^{-3}
	MC	8.49×10^{4}	59	29017	325	
+ 2 jets	ϵ	1	6.95×10^{-4}	3.42×10^{-1}	3.83×10^{-3}	9.09×10^{-7}
	n	1.73×10^{3}	1.20	5.92×10^{2}	6.63	1.57×10^{-3}
	MC	2.00×10^{4}	17	8321	139	
+ 3 jets	ϵ	1	8.50×10^{-4}	4.16×10^{-1}	6.95×10^{-3}	2.46×10^{-6}
	n	4.03×10^{2}	3.43×10^{-1}	1.68×10^{2}	2.80	9.93×10^{-4}
	MC	5.00×10^{3}	6	2408	41	
+ 4 jets	ϵ	1	1.20×10^{-3}	4.82×10^{-1}	8.21×10^{-3}	4.75×10^{-6}
	n	9.37×10^{1}	1.13×10^{-1}	4.52×10^{1}	7.69×10^{-1}	4.45×10^{-4}
Total	n	7.11×10^{5}	5.03×10^{1}	7.06×10^{4}	3.40×10^{2}	3.53×10^{-2}

Table 5.10: Results from using the cut factorisation method to estimate the contribution of W+jets background. The rows represent each sub-sample considered, with the bottom row the total W+jets contribution. Moving from left to right, the columns correspond to no cuts applied, lepton cuts, $E_{\mathrm{T}}^{\mathrm{miss}}$ cut, tag-jet cuts and the combined effect of each using the factorisation method. Number of simulated events is shown by MC, corresponding number of events normalised to luminosity used in this analysis by n and the corresponding efficiency, by ϵ.

Sample		No Cuts	Lepton Cuts	$E_{\mathrm{T}}^{\mathrm{miss}}$ cut	tag-jet cuts	Combined
Z → ττ						
+ 0 jets	MC	6.61563×10^{6}	1191	44819	1164	
	ϵ	1	1.80×10^{-4}	6.77×10^{-3}	1.76×10^{-4}	2.15×10^{-10}
	n	2.68×10^{4}	4.82	1.81×10^{2}	4.71	5.75×10^{-6}
+ 1 jets	MC	1.33×10^{6}	373	120942	330	
	ϵ	1	2.80×10^{-4}	9.06×10^{-2}	2.47×10^{-4}	6.27×10^{-9}
	n	5.42×10^{3}	1.51	4.91×10^{2}	1.34	3.40×10^{-5}
+ 2 jets	MC	4.05×10^{5}	146	67394	763	
	ϵ	1	3.61×10^{-4}	1.67×10^{-1}	1.89×10^{-3}	1.13×10^{-7}
	n	1.65×10^{3}	5.94×10^{-1}	2.74×10^{2}	3.10	1.86×10^{-4}
+ 3 jets	MC	1.10×10^{5}	56	26050	588	
	ϵ	1	5.10×10^{-4}	2.37×10^{-1}	5.35×10^{-3}	6.46×10^{-7}
	n	4.48×10^{2}	2.28×10^{-1}	1.06×10^{2}	2.40	2.90×10^{-4}
+ 4 jets	MC	3.00×10^{4}	11	8910	279	
	ϵ	1	3.67×10^{-4}	2.97×10^{-1}	9.30×10^{-3}	1.01×10^{-6}
	n	1.18×10^{2}	4.34×10^{-2}	3.51×10^{1}	1.10	1.20×10^{-4}
+ 5 jets	MC	9.99×10^{3}	4	3533	129	
	ϵ	1	4.00×10^{-4}	3.54×10^{-1}	1.29×10^{-2}	1.83×10^{-6}
	n	2.85×10^{1}	1.14×10^{-2}	1.01×10^{1}	3.68×10^{-1}	5.21×10^{-5}
Z(→ ττ)bb						
+ 0 jets	MC	1.50×10^{5}	53	5870	6	
	ϵ	1	3.54×10^{-4}	3.92×10^{-2}	4.00×10^{-5}	5.55×10^{-10}
	n	2.66×10^{2}	9.40×10^{-2}	1.04×10^{1}	1.06×10^{-2}	1.47×10^{-7}
+ 1 jets	MC	9.99×10^{4}	54	14451	82	
	ϵ	1	5.40×10^{-4}	1.45×10^{-1}	8.21×10^{-4}	6.41×10^{-8}
	n	1.01×10^{2}	5.44×10^{-2}	1.46×10^{1}	8.26×10^{-2}	6.46×10^{-6}
+ 2 jets	MC	4.00×10^{4}	21	9402	139	
	ϵ	1	5.25×10^{-4}	2.35×10^{-1}	3.48×10^{-3}	4.29×10^{-7}
	n	3.29×10^{1}	1.73×10^{-2}	7.74	1.14×10^{-1}	1.41×10^{-5}
+ 3 jets	MC	1.00×10^{4}	15	3264	72	
	ϵ	1	1.50×10^{-3}	3.27×10^{-1}	7.20×10^{-3}	3.53×10^{-6}
	n	1.58×10^{1}	2.37×10^{-2}	5.15	1.14×10^{-1}	5.57×10^{-5}
Total	n	3.49×10^{4}	7.40	1.14×10^{3}	1.33×10^{1}	7.64×10^{-4}

Table 5.11: Results from using the cut factorisation method to estimate the contribution of Z→ ττ+jets background. The rows represent each sub-sample considered, with the bottom row the total Z→ ττ+jets contribution. Moving from left to right, the columns correspond to no cuts applied, lepton cuts, $E_{\mathrm{T}}^{\mathrm{miss}}$ cut, tag-jet cuts and the combined effect of each using the factorisation method. Number of simulated events is shown by MC, corresponding number of events normalised to luminosity used in this analysis by n and the corresponding efficiency, by ϵ.

Sample		No Cuts	Lepton Cuts	$E_{\mathrm{T}}^{\mathrm{miss}}$ cut	tag-jet cuts	Combined
t channel (eν)	MC	9993	18	4913	63	
	ϵ	1	1.80×10^{-3}	4.92×10^{-1}	6.30×10^{-3}	5.58×10^{-6}
	n	7.74×10^{1}	1.39×10^{-1}	3.81×10^{1}	4.88×10^{-1}	4.32×10^{-4}
t channel (μν)	MC	1.00×10^{4}	37	5740	67	
	ϵ	1	3.70×10^{-3}	5.74×10^{-1}	6.70×10^{-3}	1.42×10^{-5}
	n	7.77×10^{1}	2.87×10^{-1}	4.46×10^{1}	5.20×10^{-1}	1.11×10^{-3}
t channel (τν)	MC	1.00×10^{4}	5	5034	57	
	ϵ	1	5.00×10^{-4}	5.03×10^{-1}	5.70×10^{-3}	1.43×10^{-6}
	n	7.71×10^{1}	3.86×10^{-2}	3.88×10^{1}	4.40×10^{-1}	1.11×10^{-4}
Wt channel	MC	1.50×10^{4}	83	5259	15	
	ϵ	1	5.54×10^{-3}	3.51×10^{-1}	1.00×10^{-3}	1.94×10^{-6}
	n	2.13×10^{2}	1.18	7.48×10^{1}	2.13×10^{-1}	4.14×10^{-4}
Total	n	4.46×10^{2}	1.65	1.96×10^{2}	1.66	2.06×10^{-3}

Table 5.12: Results from using the cut factorisation method to estimate the contribution of single top background. The rows represent each sub-sample considered, with the bottom row the total single top contribution. Moving from left to right, the columns correspond to no cuts applied, lepton cuts, $E_{\mathrm{T}}^{\mathrm{miss}}$ cut, tag-jet cuts and the combined effect of each using the factorisation method. Number of simulated events is shown by MC, corresponding number of events normalised to luminosity used in this analysis by n and the corresponding efficiency, by ϵ.

5.5 Conclusions

In this chapter basic studies have indicated that the use of single lepton triggers provides an efficient way to trigger the signal considered. A cut-based selection for its search was suggested. Two different selections were identified, in order to exploit the changing signal properties with varying Higgs mass. The low mass selection is applied for $m_H < 300\,\mathrm{GeV}/c^2$ and the high mass selection for $m_H \geq 300\,\mathrm{GeV}/c^2$. The high mass selection is able to give improved background suppression by exploiting the increasing E_T^{miss} present in the signal with increasing mass.

Reasonably good agreement between data and Monte Carlo was found for variables relating to the Higgs decay products, apart from E_T^{miss}. This difference was attributed to pile-up effects through comparison of pile-up Monte Carlo and data. As expected, the variables related to the VBF topology in the signal show an increased dependence on pile-up.

Both the low and high mass selections allow a very large suppression of background, including the main backgrounds identified, Z+jets and $t\bar{t}$. A reasonable signal efficiency of over 5 % for the $m_H = 200{,}300\,\mathrm{GeV}/c^2$ signal samples is maintained. The background contribution due to the QCD, W+jets, $Z \rightarrow \tau\tau$+jets and single top processes, here estimated by the cut factorisation method, were determined to be negligible. In the following chapter an attempt is made to optimize the selection outlined here and the mass dependence of the search is explored.

Chapter 6

Optimization of Signal Significance

In this chapter the baseline cut-based selection outlined in Chapter 5 is optimized. This is done in a two step procedure in which first the variables relating to the tag-jet selection are optimized and then variables relating to the Higgs decay products are optimized. The effect of different methods to tag the jets from the VBF remnant and different measures of the overall performance of the optimization are compared. Each optimization step is carried out at each available mass signal sample. A comparison between the baseline and optimized selection is made.

6.1 Optimization procedure

In order to establish an optimized event selection the procedure adopted was to compare different cut configurations and rank them according to a figure of merit, or significance measure by which the signal to background ratio is optimal given the inputs to the optimization. Several different definitions are typically used to define the signal significance for an analysis. The most widely used definitions are shown in Eqn. 6.1. Z_1 is valid in the Gaussian limit when signal (s) and background (b) are much greater than unity. In the Poisson limit, this definition is approximate and only valid when $s \ll b$ and s and b are known to a high accuracy. In this analysis it has been shown that $s \ll b$ and therefore it is expected that this definition is suitable. However, Z_2 [84] is a more general definition of Z_1 in that it is valid when $s \gg b$ in the Poisson limit. In the following both measures are compared.

$$Z_1 = \frac{s}{\sqrt{b}} \qquad Z_2 = \sqrt{2((s+b)ln(1+\frac{s}{b}) - s)} \tag{6.1}$$

The optimization procedure takes into account correlations between the cuts being optimized by considering each possible combination of cuts, i.e. performing an

optimization in an n dimensional cut-space, where n is the number of different cut variables being optimized. Due to the large CPU time required to perform optimization done in this way (number of combinations = Number of Cuts per Variable^n) a method other than considering each combination of cuts was developed and was used to derive results shown in this chapter. In this method, the cut combinations are ordered in increasing tightness and information of what the previous cut combination passed is used to identify which subsequent cut should be considered. Through this method, intermediate combinations are assigned as being passed or failed depending on whether the cuts are looser or tighter and this allows for the computation time of evaluating the total cut-space to be drastically reduced. Results using this method were compared to the more time consuming method of considering each combination in turn for a variety of different situations, where the number of variables and the number of cuts per variable was made different. Agreement between the two methods was found in all cases.

The identification of an optimal set of cuts using the optimization procedure outlined is done using an independent Monte Carlo (training) sample with the same size as that used to produce the results of the baseline selection (testing sample) shown in Chapter 5. Final results using the optimized cuts will be quoted using the testing sample so that a direct comparison of each selection can be made. All processes are used in the optimization apart from those contributing to QCD and W+jets because the available statistics is too small. However, as shown in Section 5.4 these background are expected to have negligible contribution to the overall background after all cuts are applied.

6.2 Optimization of tag-jet selection

The variables considered in the optimization of variables related to the VBF remnants included the leading tag-jet p_{T}, the requirement that the tag-jets be in opposite hemispheres ($\eta \times \eta$), the $\Delta\eta$ between the tag-jets, the mass of the tag-jets and the p_{T} of non tag-jets within $|\eta| < 3.2$. The ranges and step sizes considered for each of the variables considered is shown below. In each case the possibility of applying no-cut was also tested.

- leading tag-jet $p_{\mathrm{T}} >$ 20-65 GeV/c in steps of 5 GeV/c.
- $\eta \times \eta < 0$.
- $\Delta\eta >$ 3.2-4.8 in steps of 0.2.
- p_{T} (non-tag-jets) 5-95 GeV/c in steps of 10 GeV/c.

- mass (tag-jets) 300-750 GeV/c^2 in steps of 50 GeV/c^2.

In order to ensure that the background events used in the optimization process are relevant to the search in the VBF $\mathrm{H} \to \mathrm{ZZ} \to ll\nu\nu$ channel, events are required to pass the lepton selection, $E_{\mathrm{T}}^{\mathrm{miss}}$ and Z mass selections from the baseline selection. Attempts were made to perform the optimization loosening these cuts further to maximise statistics available. Typically however, these optimizations tended to select very hard optimal cuts far outside what would generally be used in a VBF analysis. This is because of the overwhelming Z+jets background which has to be dealt with, in particular, once $E_{\mathrm{T}}^{\mathrm{miss}}$ is loosened to below 30 GeV (as shown in Figure 5.6(b)). Because of these problems the low mass selection baseline cuts relating to the Higgs decay products were adopted in the optimizations performed of the VBF related variables and used to produce the results shown subsequently. However, requiring this selection meant that there were very limited statistics with which to perform the optimization. In an attempt to avoid problems associated with allowing the optimization to explore regions of cut space that are not well understood due to lack of statistics, for each combination of cuts, when a certain sample runs out of Monte Carlo statistics, then the contribution from this background is estimated by assuming one surviving Monte Carlo event. The total background found using this method is then required to be less than the sum of the total background calculated without this replacement and its associated error, in order for the combination of cuts to be considered as valid.

Example distributions showing the maximum significance as a function of cut value, for the $\mathrm{m_H} = 200\,\mathrm{GeV}/c^2$ signal sample optimization are shown in Figure 6.1, including a comparison of $\mathrm{Z_1}$ (shown in green) and $\mathrm{Z_2}$ (shown in blue) significance measures. No large difference between each significance measure is observed to exist. Using the inputs to the optimization described, the most discriminating variable is the $\Delta\eta$ cut. The results from repeating the optimization of VBF related variables at each mass point considered is shown in Figure 6.2, where the optimal cuts are plotted as a function of Higgs mass for each of the variables considered. The effect of using each significance measure is also shown, $\mathrm{Z_1}$ by the filled green squares and $\mathrm{Z_2}$ by the open blue squares. Using the inputs described, the optimal cut positions are fairly independent of Higgs mass. Each significance measure is shown to select the same optimal cut positions.

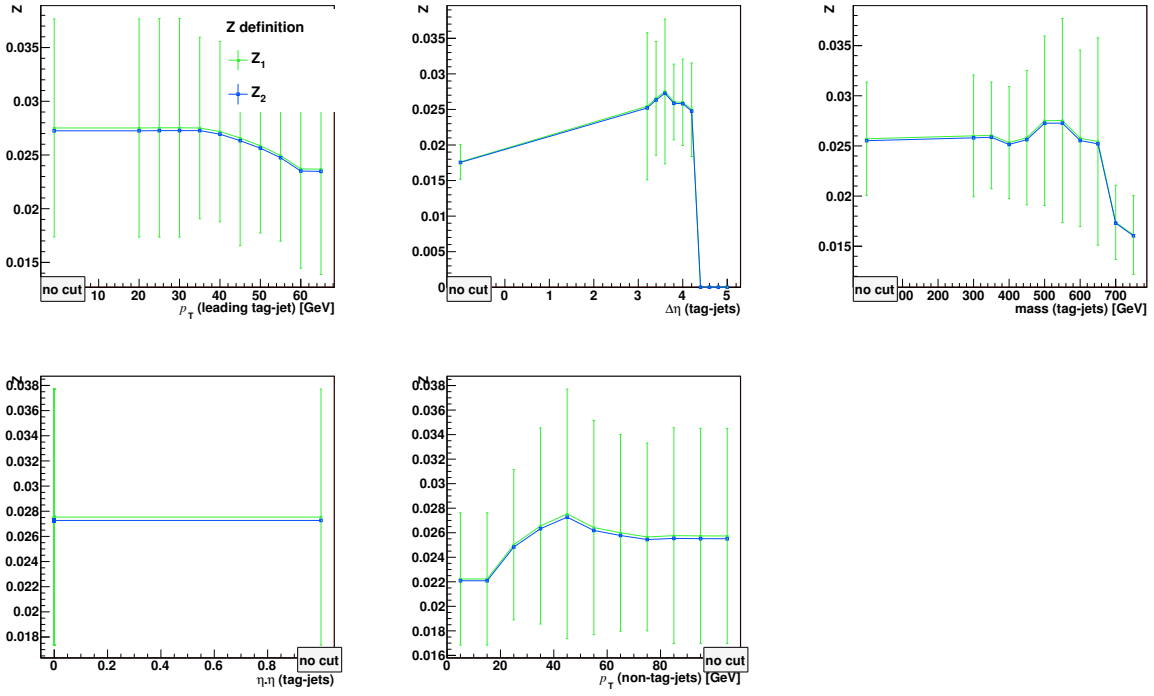


Figure 6.1: Distributions of maximum significance as a function of cut value for the leading tag-jet p_{T}, $\Delta\eta$ (tag-jets), mass (tag-jets), $\eta \times \eta$ (tag-jets) and p_{T} (non tag-jets) variables. Signal corresponds to the $\mathrm{m_H} = 200\,\mathrm{GeV}/c^2$ sample. Included are distributions from using the significance measures $\mathrm{Z_1}$ and $\mathrm{Z_2}$.

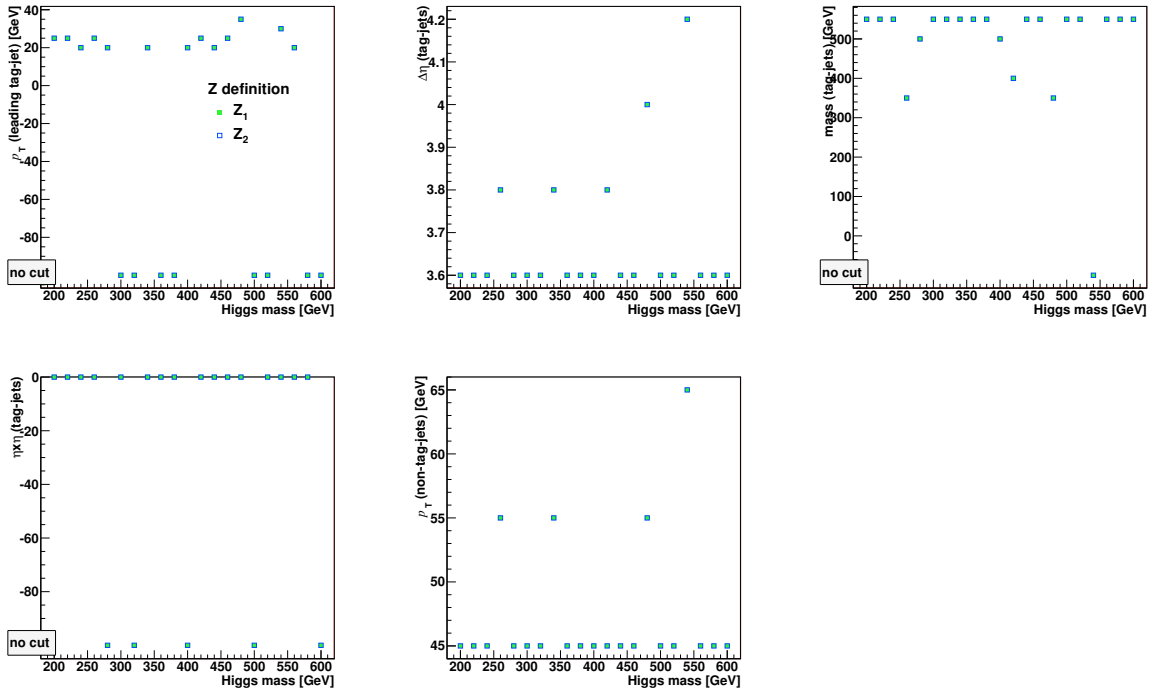


Figure 6.2: Distributions of optimized cut positions as a function of Higgs mass for the leading tag-jet p_{T}, $\Delta\eta$ (tag-jets), mass (tag-jets), $\eta \times \eta$ (tag-jets) and p_{T} (non tag-jets) variables. Included are the results from using the significance measures $\mathrm{Z_1}$ and $\mathrm{Z_2}$.

6.2.1 Comparison of methods to tag the jets from the remnants of the VBF process

As discussed in Section 5.3 the tag-jets in VBF analyses are most commonly identified using the two highest p_{T} jets. In the following a comparison is made between this approach and two other methods: i) considering the tag-jets as the two jets with the highest invariant mass and ii) the two jets with the largest separation in η. The best performing method is then subsequently compared to the more conventional one of using the highest p_{T} jets by re-performing the optimizations of the previous section.

In order to make a comparison of these different methods, the minimum combined ΔR between the tag-jets identified with each method and the quarks generating the VBF topology (subsequently referred to as truth VBF quark's, required to have $p_{\mathrm{T}} > 20$ GeV/c and $|\eta| < 4.5$) was found. This study was done using the $\mathrm{m_H} = 200\,\mathrm{GeV}/c^2$ signal sample. Reconstructed jets satisfying the criteria to be used in the analysis, as detailed in Section 4.4.1 were used. The same was done for the best matched pair of reconstructed jets. In order to establish a matching efficiency, the ΔR between each truth VBF quark and its matched reconstructed jet was required to be less than 0.2. Using this definition the matching efficiencies shown in Table 6.1 were derived. The best matched reconstructed jets are found to satisfy this criteria for approximately 70 % of events in truth $\mathrm{Z} \to ee$ events and around 75 % of events in truth $\mathrm{Z} \to \mu\mu$ events for the $\mathrm{m_H} = 200\,\mathrm{GeV}/c^2$ signal sample. Selecting tag-jets based on the highest p_{T} and mass jets was shown to give a similar performance in truth $\mathrm{Z} \to \mu\mu$ events, whereby an efficiency of around 90 % with respect to the best matched scenario is achieved in each method. However, in truth $\mathrm{Z} \to ee$ events the performance of selecting tag-jets based on highest mass gives an improvement of over 10 % compared to using the highest p_{T} jets. This trend is also indicated in Figure 6.3 which compares the p_{T} difference between reconstructed and truth VBF quarks as a function of truth VBF quark p_{T} and $|\eta|$ for truth $\mathrm{Z} \to \mu\mu/ee$ events in the $\mathrm{m_H} = 200\,\mathrm{GeV}/c^2$ signal sample. The performance of using the jets with the largest separation in η also shows an improvement over the method using highest p_{T} jets in truth $\mathrm{Z} \to ee$ events but is worse in truth $\mathrm{Z} \to \mu\mu$ events. However, its performance is worse compared to the method using the highest mass tag-jets and so is not discussed further.

In order to try and better establish if selecting tag-jets based on highest mass jets in the event provides an improvement over using the highest p_{T} jets, the two methods were compared by re-running the optimization procedure as detailed earlier in this section using the tag-jets identified in the different methods. This is expected to provide a much more robust comparison of methods to tag the jets because it takes

Method	Z → *ee* channel	Z → $\mu\mu$ channel
Best Matched	69	74
Highest p_T	48	67
Highest Mass	59	67
Largest η separation	55	61

Table 6.1: Comparison of methods to tag the jets from the VBF remnants based on comparison of the combined ΔR between the reconstructed tag-jets and the Monte Carlo truth VBF quarks. The values quoted correspond to the percentage of events with $\Delta R < 0.2$. Uncertainty on values $\approx 1\,\%$.

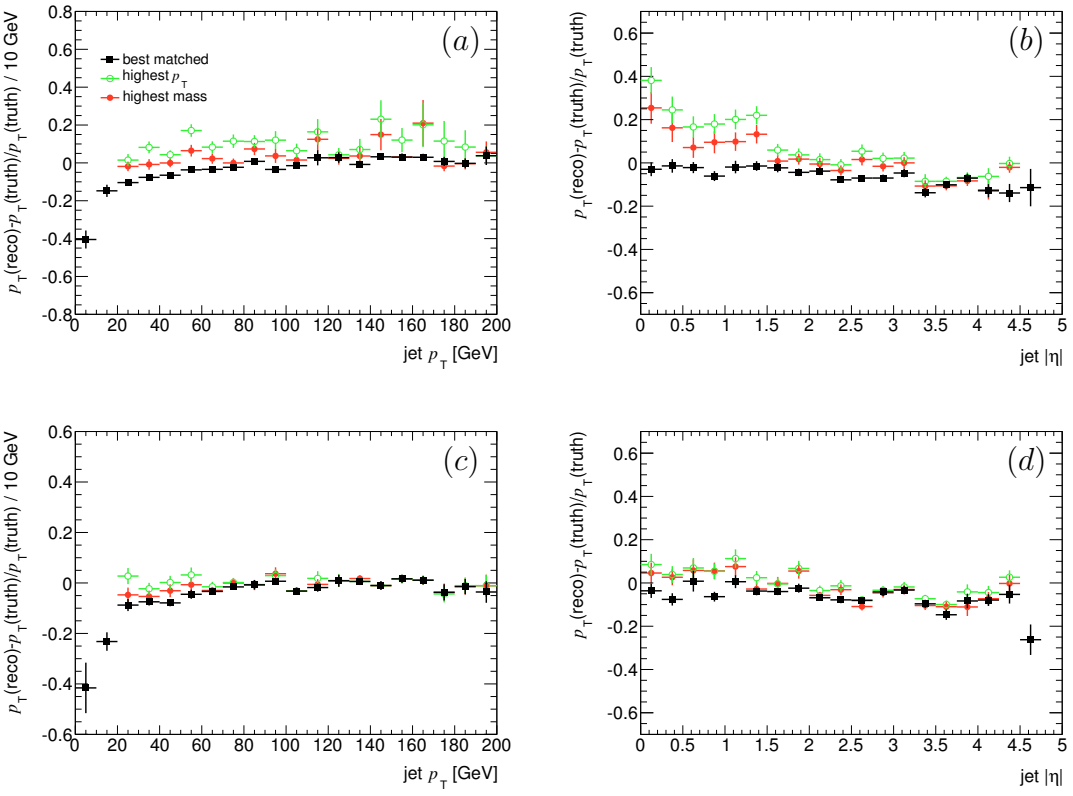


Figure 6.3: Profiles of p_{T}(reco)-p_{T}(truth)/p_{T}(truth) as a function of truth VBF quark η and p_{T} for the (a,b) Z → *ee* and (c,d) Z → $\mu\mu$ channels.(reco/truth) refers to the reconstructed jet/truth VBF quark respectively.

into account the signal and the background and chooses the method based on which provides better signal over background discrimination. The significance measure Z_1 was used. The result of this optimization for the $m_H = 200\,\mathrm{GeV}/c^2$ signal sample is shown in Figure 6.4. Each method appears to perform similarly, although the method using the highest p_T jets as the tag-jets yields a slightly higher significance. A similar trend was observed for the other signal mass points considered, as shown in Figure 6.5, and as such the method using highest p_T jets is adopted to obtain the optimized results.

Considering the little mass dependence of the optimizations performed, a fixed set of cuts relating to the VBF topology was applied across the mass points considered. The value of the cuts used was found by taking the average of the cut values found as a function of Higgs mass. The optimal cut values are leading tag-jet $p_T > 25\,\mathrm{GeV}/c$, requiring the tag-jets are in opposite hemispheres, $\Delta\eta > 3.6$, non tag-jet $p_T > 45\,\mathrm{GeV}/c$ and mass tag-jets $> 450\,\mathrm{GeV}/c^2$.

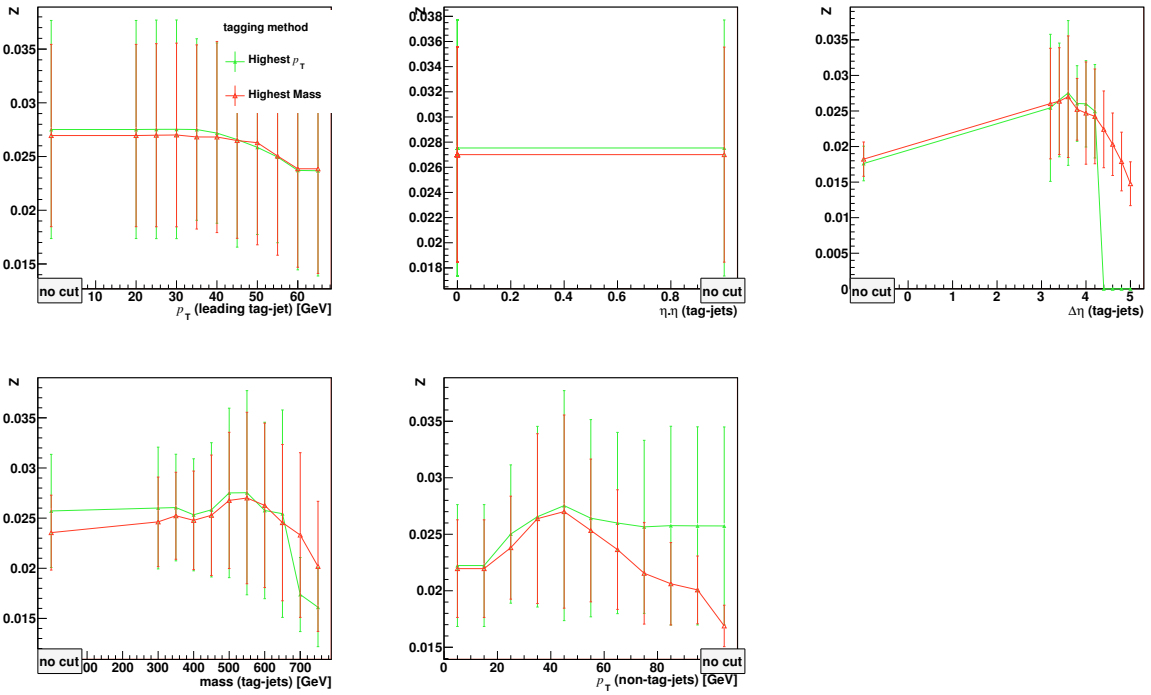


Figure 6.4: Distributions of maximum significance as a function of cut value for the leading tag-jet p_T, $\Delta\eta$ (tag-jets), mass (tag-jets), $\eta \times \eta$ (tag-jets) and p_T (non tag-jets) variables. Signal corresponds to the $m_H = 200\,\mathrm{GeV}/c^2$ sample. Included are distributions from using the highest p_T jets as the tag-jets (green filled triangular markers) and the highest mass jets as the tag-jets (open red triangular markers).

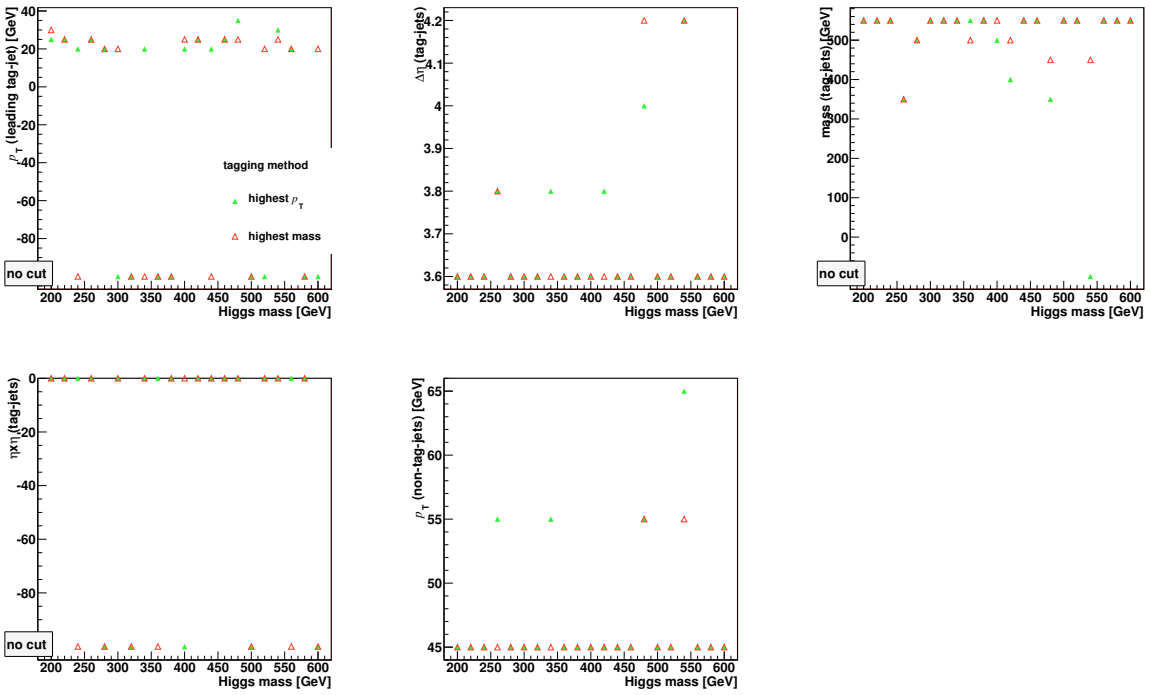


Figure 6.5: Distributions of optimized cut positions as a function of Higgs mass for the leading tag-jet p_{T}, $\Delta\eta$ (tag-jets), mass (tag-jets), $\eta \times \eta$ (tag-jets) and p_{T} (non tag-jets) variables. Included are the results from using the highest p_{T} jets as the tag-jets (green filled triangular markers) and the highest mass jets as the tag-jets (open red triangular markers).

6.3 Optimization of variables relating to Higgs decay products

The variables relating to the Higgs decay products were optimized using the same method as used to optimize the variables relating to the Higgs decay products. This was done using the Z_1 and Z_2 significance measures with cuts on the variables relating VBF topology fixed at the optimized values. The variables optimized, the cut range investigated and the step sizes used were

- Lepton $p_T >$, 20-70 GeV/c in steps of 5 GeV/c.
- (di-lepton mass-m_Z) $<$, mass 20-70 GeV/c^2 in steps of 5 GeV/c^2.
- $E_T^{miss} >$, 20-270 in steps of 25 GeV, then (selected-25)-(selected+25) GeV in steps of 5 GeV.

where in the case of E_T^{miss}, because of the large potential range of optimal cuts (taking into account varying Higgs mass), the optimization was done in two steps. First E_T^{miss} in the range 20-270 GeV in 25 GeV steps was considered and then which ever cut was found to be the optimum (selected), the range (selected-25)-(selected+25) was then considered in steps of 5 GeV.

Some examples of the optimization results for the $m_H = 200\,\text{GeV}/c^2$ signal sample are shown in Figure 6.6. As with the optimization relating to the VBF topology, the results showed little dependence on significance measure. The optimal cuts for the low and high mass selections were obtained by averaging the optimal cuts found at each mass point in a similar way to when the optimization of the VBF related variables was done. For the low mass selection, the masses between m_H = 200-280 GeV/c^2 were used and the optimal cuts identified as lepton $p_T > 20\,\text{GeV}/c$, di-lepton mass window $|m_{ll} - m_Z| < 15\,\text{GeV}/c^2$ and $E_T^{miss} > 40\,\text{GeV}$. For the high mass selection $m_H \geq 300\,\text{GeV}/c^2$ samples were used and the optimal cuts found to be lepton $p_T > 20\,\text{GeV}/c$, di-lepton mass window $|m_{ll} - m_Z| < 15\,\text{GeV}/c^2$ and $E_T^{miss} > 50\,\text{GeV}$.

6.4 Results of optimization

In summary, the optimal cuts adopted for the low and high mass selections are

- Lepton Selection: Two electrons or muons passing requirements outlined in Sections 4.2.1 and 4.3.1. No other leptons of any other type. Same flavour opposite sign requirement made in muon channel.
- Di-lepton mass window: $|m_{ll} - m_Z| < 15\,\text{GeV}/c^2$.

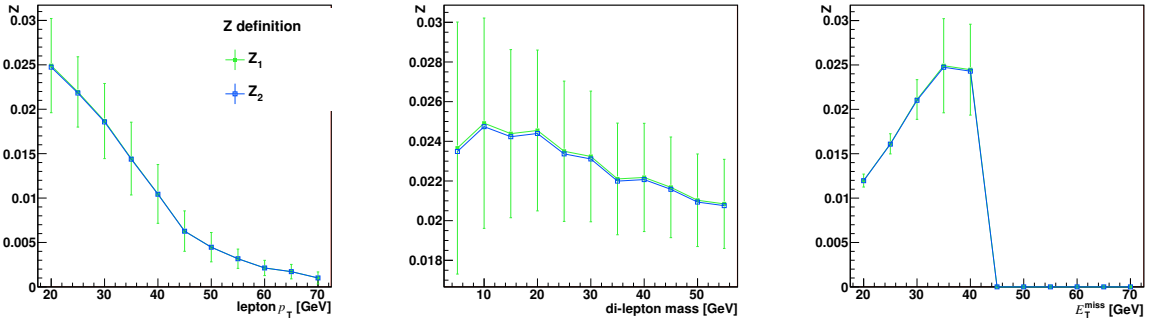


Figure 6.6: Distributions of maximum significance as a function of cut value for the lepton p_{T}, di-lepton mass and $E_{\mathrm{T}}^{\mathrm{miss}}$ variables. Signal corresponds to the $\mathrm{m_H} = 200\,\mathrm{GeV}/c^2$ sample. Included are distributions from using the significance measures $\mathrm{Z_1}$ and $\mathrm{Z_2}$.

- $E_{\mathrm{T}}^{\mathrm{miss}}$: $> 40(50)\,\mathrm{GeV}$.
- $\Delta\phi$: Between leptons $>0.5(<2)$ rad.
- b-jet veto: Reject any event with a jet with SV0 weight > 5.72.
- m_{T}: no cut$(> 200)\,\mathrm{GeV}/c^2$.
- $\mathrm{N_{jets}}$: ≥ 2 jets passing selection outlined in Section 4.4.1.
- leading tag-jet p_{T}: $> 25\,\mathrm{GeV}/c$
- $\eta \times \eta$: Product of tag-jets absolute η must be less than 0, i.e. tag-jets must be in opposite hemispheres.
- $\Delta\eta$: > 3.6(no cut).
- CJV: Reject events with non tag-jets with $p_{\mathrm{T}} > 45\,\mathrm{GeV}/c$ and $|\eta| < 2.5$ (no cut).
- mass tag-jets: $> 450\,\mathrm{GeV}/c^2$(no cut).

Where there is a difference between the selections, the high mass selection is quoted in brackets.

The results of applying this optimized selection to the testing sample are shown in Tables 6.2-6.4, which show the number of simulated events surviving the optimal cuts and the number of normalised events expected to pass the low(high) mass selections for the $\mathrm{m_H} = 200(300)\,\mathrm{GeV}/c^2$ signal samples respectively. The method for estimating the contribution of W and QCD backgrounds discussed in Chapter 5 applies to the optimal set of cuts because the cuts used to estimate their effect are tighter than those in the baseline. Applying the optimal cuts, the contribution of

W and QCD backgrounds estimated using cut factorization is shown to be within the statistical error on the background.

In the low mass selection for the $m_H = 200\,\mathrm{GeV}/c^2$ signal, using the optimized selection an improvement in signal significance of just under 10 % is observed. A similar effect is seen for the high mass selection for the $m_H = 300\,\mathrm{GeV}/c^2$ signal. Signal efficiencies in both cases are maintained above 5 %.

Optimized Low Mass Selection (Z → $\mu\mu/ee$ channels combined)						
Sample	lepton selection	di-lepton mass window	$E_{\mathrm{T}}^{\mathrm{miss}}$	$\Delta\phi$	b-jet veto	m_{T}
VBF	1142	1085	642	617	602	602
GF	4139	3933	1692	1657	1635	1635
$t\bar{t}$	11885	2540	1849	1739	603	603
WW	21921	5033	2288	2192	2171	2171
ZZ	38523	35883	14312	13502	13370	13370
WZ	50166	44770	9750	9194	9042	9042
Z+jets	2778532	2523711	2566	2207	2095	2095
W+jets	324	83	19	18	17	17
Wbb+jets	140	38	11	10	8	8
Zbb+jets	113884	104576	1365	1183	600	600
QCD	555	53	3	3	3	3
single top	120	20	13	13	5	5
Z→ $\tau\tau$+jets	7159	381	27	27	27	27

Optimized Low Mass Selection (Z → $\mu\mu/ee$ channels combined)							
Sample	Initial	$N_{\mathrm{jets}} \geq 2$	leading tag-jet p_T	$\eta \times \eta$	$\Delta\eta$	CJV	mass tag-jets
VBF	602	470	470	386	244	238	200
GF	1635	467	456	214	68	66	25
$t\bar{t}$	603	509	507	215	34	28	14
WW	2171	304	292	145	36	33	10
ZZ	13370	1947	1880	805	168	168	57
WZ	9042	2876	2779	1172	231	218	96
Z+jets	2095	1065	1063	473	37	31	16
W+jets	17	5	3	3	0	0	0
Wbb+jets	8	3	3	1	0	0	0
Zbb+jets	600	436	432	178	11	9	6
QCD	3	1	1	0	0	0	0
single top	5	2	2	1	0	0	0
Z→ $\tau\tau$+jets	27	9	9	3	0	0	0

Optimized High Mass Selection (Z → $\mu\mu/ee$ channels combined)						
Sample	lepton selection	di-lepton mass window	$E_{\mathrm{T}}^{\mathrm{miss}}$	$\Delta\phi$	b-jet veto	m_{T}
VBF	1295	1212	1107	949	922	899
GF	4309	4005	3638	3191	3166	3106
$t\bar{t}$	11885	2540	1549	748	258	208
WW	21921	5033	1508	856	848	803
ZZ	38523	35883	11025	6880	6808	6659
WZ	50166	44770	7010	4077	3999	3809
Z+jets	2778532	2523711	1077	714	658	611
W+jets	324	83	13	6	6	6
Wbb+jets	140	38	3	0	0	0
Zbb+jets	113884	104576	676	517	226	218
QCD	555	53	2	0	0	0
single top	120	20	10	4	0	0
Z→ $\tau\tau$+jets	7159	381	17	2	2	1

Optimized High Mass Selection (Z → $\mu\mu/ee$ channels combined)							
Sample	Initial	$N_{\mathrm{jets}} \geq 2$	leading tag-jet p_T	$\eta \times \eta$	$\Delta\eta$	CJV	mass tag-jets
VBF	899	684	680	545	402	402	402
GF	3106	822	793	378	107	107	107
$t\bar{t}$	208	173	171	77	13	13	13
WW	803	102	97	47	16	16	16
ZZ	6659	932	900	379	74	74	74
WZ	3809	1258	1219	505	88	88	88
Z+jets	611	443	443	196	13	13	13
W+jets	6	2	1	1	0	0	0
Wbb+jets	0	0	0	0	0	0	0
Zbb+jets	218	181	181	77	5	5	5
QCD	0	0	0	0	0	0	0
single top	0	0	0	0	0	0	0
Z→ $\tau\tau$+jets	1	1	1	0	0	0	0

Table 6.2: Number of simulated events passing the optimized selection for the low and high mass selections. Included are the combined statistics for the Z → $\mu\mu/ee$ channels. The signal corresponds to the $m_{\mathrm{H}} = 200(300)\,\mathrm{GeV}/c^2$ sample for the low(high) mass selection.

Optimized Low Mass Selection (Z → $\mu\mu/ee$ channels combined)						
Sample	lepton selection	di-lepton mass window	$E_{\mathrm{T}}^{\mathrm{miss}}$	$\Delta\phi$	b-jet veto	m_{T}
VBF	0.08±0.002 (0.35)	0.07±0.002 (0.33)	0.04±0.002 (0.20)	0.04±0.002 (0.19)	0.04±0.002 (0.18)	0.04±0.002 (0.18)
GF	0.65±0.01 (0.36)	0.61±0.0098 (0.34)	0.26±0.0064 (0.15)	0.26±0.0064 (0.14)	0.26±0.0063 (0.14)	0.26±0.0063 (0.14)
$t\bar{t}$	72.78±0.67 (2.4×10^{-2})	15.55±0.31 (5.1×10^{-3})	11.32±0.26 (3.7×10^{-3})	10.65±0.26 (3.5×10^{-3})	3.69±0.15 (1.2×10^{-3})	3.69±0.15 (1.2×10^{-3})
WW	15.65±0.11 (9.6×10^{-2})	3.59±0.05 (2.2×10^{-2})	1.64±0.03 (1.0×10^{-2})	1.57±0.03 (9.6×10^{-3})	1.55±0.03 (9.5×10^{-3})	1.55±0.03 (9.5×10^{-3})
ZZ	11.59±0.12 (0.30)	11.26±0.12 (0.29)	1.55±0.02 (4.0×10^{-2})	1.45±0.02 (3.7×10^{-2})	1.41±0.02 (3.6×10^{-2})	1.41±0.02 (3.6×10^{-2})
WZ	15.78±0.11 (0.12)	14.77±0.11 (0.11)	1.74±0.03 (1.3×10^{-2})	1.63±0.03 (1.2×10^{-2})	1.59±0.03 (1.2×10^{-2})	1.59±0.03 (1.2×10^{-2})
Z+jets	22695±14 (0.33)	20642±13 (0.30)	20.57±0.41 (3.0×10^{-4})	17.71±0.38 (2.6×10^{-4})	16.82±0.37 (2.4×10^{-4})	16.82±0.37 (2.4×10^{-4})
W+jets	79.03±5.14 (1.2×10^{-4})	21.40±2.98 (3.1×10^{-5})	4.73±1.78 (6.9×10^{-6})	4.34±1.72 (6.3×10^{-6})	4.18±1.68 (6.1×10^{-6})	4.18±1.68 (6.1×10^{-6})
Wbb+jets	5.68±0.52 (2.1×10^{-4})	1.51±0.35 (5.7×10^{-5})	0.42±0.25 (1.6×10^{-5})	0.38±0.24 (1.4×10^{-5})	0.31±0.21 (1.1×10^{-5})	0.31±0.21 (1.1×10^{-5})
Zbb+jets	320.1±1 (0.39)	294.5±1 (0.35)	3.27±0.09 (3.9×10^{-3})	2.83±0.09 (3.4×10^{-3})	1.44±0.06 (1.7×10^{-3})	1.44±0.06 (1.7×10^{-3})
QCD	1039±327 (2.9×10^{-9})	155.5±237.7 (4.4×10^{-10})	2.12±2.88 (6.0×10^{-12})	2.12±2.88 (6.0×10^{-12})	2.12±2.88 (6.0×10^{-12})	2.12±2.88 (6.0×10^{-12})
single top	3.16±0.30 (7.1×10^{-3})	0.52±0.13 (1.2×10^{-3})	0.32±0.13 (7.2×10^{-4})	0.32±0.13 (7.2×10^{-4})	0.11±0.10 (2.6×10^{-4})	0.11±0.10 (2.6×10^{-4})
Z→ ττ+jets	56.16±0.67 (1.6×10^{-3})	3.01±0.16 (8.6×10^{-5})	0.21±0.05 (6.2×10^{-6})	0.21±0.05 (6.2×10^{-6})	0.21±0.05 (6.2×10^{-6})	0.21±0.05 (6.2×10^{-6})
Total	24314±328 (6.9×10^{-8})	21164±238 (6.0×10^{-8})	47.89±3.44 (1.4×10^{-10})	43.21±3.40 (1.2×10^{-10})	33.45±3.37 (9.5×10^{-11})	33.45±3.37 (9.5×10^{-11})
Data	23633	20484	45	42	36	36
$s/\sqrt{b}$	0.0005±0.00002	0.0005±0.00002	0.006±0.0003	0.006±0.0003	0.007±0.0003	0.007±0.0003

Optimized Low Mass Selection (Z → $\mu\mu/ee$ channels combined)								
bkgd	Initial	$N_{\mathrm{jets}} \geq 2$	leading tag-jet p_T	$\eta \times \eta$	$\Delta\eta$	CJV	mass tag-jets	1Event(Total)
VBF	0.04±0.002 (0.18)	0.03±0.001 (0.14)	0.03±0.001 (0.14)	0.03±0.001 (0.12)	0.02±0.001 (7.5×10^{-2})	0.02±0.001 (7.3×10^{-2})	0.01±0.001 (6.1×10^{-2})	0.01±0.001
GF	0.26±0.0063 (0.14)	0.07±0.003 (4.0×10^{-2})	0.07±0.003 (3.9×10^{-2})	0.03±0.002 (1.8×10^{-2})	0.01±0.001 (5.8×10^{-3})	0.01±0.001 (5.7×10^{-3})	0.004±0.0008 (2.1×10^{-3})	0.004±0.0008
$t\bar{t}$	3.69±0.15 (1.2×10^{-3})	3.12±0.14 (1.0×10^{-3})	3.10±0.14 (1.0×10^{-3})	1.32±0.09 (4.3×10^{-4})	0.21±0.04 (6.8×10^{-5})	0.17±0.03 (5.7×10^{-5})	0.08±0.02 (2.7×10^{-5})	0.08±0.02
WW	1.55±0.03 (9.5×10^{-3})	0.22±0.01 (1.3×10^{-3})	0.21±0.01 (1.3×10^{-3})	0.11±0.01 (6.4×10^{-4})	0.03±0.006 (1.6×10^{-4})	0.02±0.006 (1.5×10^{-4})	0.007±0.004 (4.4×10^{-5})	0.01±0.006
ZZ	1.41±0.02 (3.6×10^{-2})	0.23±0.009 (5.8×10^{-3})	0.22±0.009 (5.6×10^{-3})	0.09±0.007 (2.3×10^{-3})	0.02±0.001 (4.1×10^{-4})	0.02±0.001 (4.1×10^{-4})	0.0055±0.008 (1.4×10^{-4})	0.008±0.004
WZ	1.59±0.03 (1.2×10^{-2})	0.55±0.02 (4.1×10^{-3})	0.53±0.02 (3.9×10^{-3})	0.22±0.01 (1.6×10^{-3})	0.04±0.005 (3.1×10^{-4})	0.04±0.005 (2.9×10^{-4})	0.02±0.002 (1.3×10^{-4})	0.03±0.01
Z+jets	16.82±0.37 (2.4×10^{-4})	8.44±0.26 (1.2×10^{-4})	8.42±0.26 (1.2×10^{-4})	3.72±0.17 (5.4×10^{-5})	0.29±0.08 (4.2×10^{-6})	0.25±0.07 (3.6×10^{-6})	0.12±0.06 (1.8×10^{-6})	0.15±0.07
W+jets	4.18±1.68 (6.1×10^{-6})	0.77±0.70 (1.1×10^{-6})	0.46±0.56 (6.7×10^{-7})	0.46±0.56 (6.7×10^{-7})	-	-	-	3.83±2.24
Wbb+jets	0.31±0.21 (1.1×10^{-5})	0.11±0.13 (4.1×10^{-6})	0.11±0.13 (4.1×10^{-6})	0.04±0.09 (1.3×10^{-6})	-	-	-	0.35±0.27
Zbb+jets	1.44±0.06 (1.7×10^{-3})	0.96±0.05 (1.2×10^{-3})	0.95±0.05 (1.1×10^{-3})	0.39±0.03 (4.7×10^{-4})	0.02±0.01 (2.8×10^{-5})	0.02±0.01 (2.1×10^{-5})	0.01±0.008 (1.2×10^{-5})	0.03±0.02
QCD	2.12±2.88 (6.0×10^{-12})	0.002±0.004 (4.5×10^{-15})	0.002±0.004 (4.5×10^{-15})	-	-	-	-	505119±1078932
single top	0.11±0.10 (2.6×10^{-4})	0.06±0.07 (1.2×10^{-4})	0.06±0.07 (1.2×10^{-4})	0.03±0.07 (6.4×10^{-5})	-	-	-	0.07±0.09
Z→ ττ+jets	0.21±0.05 (6.2×10^{-6})	0.06±0.04 (1.9×10^{-6})	0.06±0.04 (1.9×10^{-6})	0.02±0.03 (6.9×10^{-7})	-	-	-	0.06±0.05
Total	33.45±3.37 (9.47×10^{-11})	14.51±0.78 (4.11×10^{-11})	14.13±0.66 (4.00×10^{-11})	6.39±0.61	0.61±0.09 (1.72×10^{-12})	0.52±0.08 (1.47×10^{-12})	0.25±0.06 (6.98×10^{-13})	0.30±0.07
Data	36	20	20	4	0	0	0	
$s/\sqrt{b}$	0.007±0.0003	0.008±0.0004	0.008±0.0004	0.01±0.0006	0.021±0.002	0.022±0.003	0.027±0.007	0.025±0.02

Table 6.3: Expected number of simulated events passing optimized low mass selection given an integrated luminosity of $33.4\,\mathrm{pb}^{-1}$, for the combined Z → $\mu\mu/ee$ channels. Efficiencies are given in brackets. The signal corresponds to the $\mathrm{m_H} = 200\,\mathrm{GeV}/c^2$ sample.

Optimized High Mass Selection (Z → $\mu\mu/ee$ channels combined)						
Sample	lepton selection	di-lepton mass window	$E_{\mathrm{T}}^{\mathrm{miss}}$	$\Delta\phi$	b-jet veto	m_{T}
VBF	0.05±0.001 (0.38)	0.04±0.001 (0.36)	0.04±0.001 (0.33)	0.04±0.001 (0.28)	0.03±0.001 (0.27)	0.03±0.001 (0.27)
GF	0.38±0.0057 (0.38)	0.35±0.0055 (0.35)	0.32±0.0053 (0.32)	0.28±0.0049 (0.28)	0.28±0.0049 (0.28)	0.27±0.0049 (0.27)
$t\bar{t}$	72.78±0.67 (2.4×10^{-2})	15.55±0.31 (5.1×10^{-3})	9.48±0.24 (3.1×10^{-3})	4.58±0.17 (1.5×10^{-3})	1.58±0.10 (5.2×10^{-4})	1.28±0.09 (4.2×10^{-4})
WW	15.65±0.11 (9.6×10^{-2})	3.59±0.05 (2.2×10^{-2})	1.08±0.03 (6.6×10^{-3})	0.61±0.02 (3.7×10^{-3})	0.60±0.02 (3.7×10^{-3})	0.57±0.02 (3.5×10^{-3})
ZZ	11.59±0.12 (0.30)	11.26±0.12 (0.29)	1.17±0.01 (3.0×10^{-2})	0.75±0.01 (1.9×10^{-2})	0.73±0.01 (1.9×10^{-2})	0.71±0.01 (1.8×10^{-2})
WZ	15.78±0.11 (0.12)	14.77±0.11 (0.11)	1.23±0.02 (9.2×10^{-3})	0.74±0.02 (5.5×10^{-3})	0.72±0.02 (5.4×10^{-3})	0.69±0.02 (5.1×10^{-3})
Z+jets	22695±14 (0.33)	20642±13 (0.30)	8.63±0.26 (1.2×10^{-4})	5.71±0.21 (8.3×10^{-5})	5.26±0.21 (7.6×10^{-5})	4.89±0.20 (7.1×10^{-5})
W+jets	79.03±5.14 (1.2×10^{-4})	21.40±2.98 (3.1×10^{-5})	3.10±1.52 (4.5×10^{-6})	1.63±1.30 (2.4×10^{-6})	1.63±1.30 (2.4×10^{-6})	1.63±1.30 (2.4×10^{-6})
Wbb+jets	5.68±0.52 (2.1×10^{-4})	1.51±0.35 (5.7×10^{-5})	0.11±0.16 (4.3×10^{-6})	-	-	-
Zbb+jets	320.1±1 (0.39)	294.5±1 (0.35)	1.61±0.07 (1.9×10^{-3})	1.24±0.06 (1.5×10^{-3})	0.53±0.04 (6.4×10^{-4})	0.52±0.04 (6.2×10^{-4})
QCD	1039±327 (2.9×10^{-9})	155.5±237.7 (4.4×10^{-10})	1.04±2.51 (2.9×10^{-12})	-	-	-
single top	3.16±0.30 (7.1×10^{-3})	0.52±0.13 (1.2×10^{-3})	0.24±0.12 (5.3×10^{-4})	0.10±0.09 (2.2×10^{-4})	-	-
Z→ ττ+jets	56.16±0.67 (1.6×10^{-3})	3.01±0.16 (8.6×10^{-5})	0.14±0.05 (3.9×10^{-6})	0.01±0.02 (2.9×10^{-7})	0.01±0.02 (2.9×10^{-7})	0.002±0.005 (5.5×10^{-8})
Total	24314±328 (6.9×10^{-8})	21164±238 (6.0×10^{-8})	27.84±2.96 (7.9×10^{-11})	15.35±1.33 (4.3×10^{-11})	11.07±1.32 (3.1×10^{-11})	10.28±1.32 (2.9×10^{-11})
Data	23633	20484	17	10	10	9
$s/\sqrt{b}$	0.0003±0.000009	0.0003±0.000009	0.008±0.0002	0.009±0.0003	0.01±0.0004	0.01±0.0004

Optimized High Mass Selection (Z → $\mu\mu/ee$ channels combined)								
Sample	Initial	$N_{\mathrm{jets}} \geq 2$	leading tag-jet p_T	$\eta \times \eta$	$\Delta\eta$	CJV	mass tag-jets	1Event(Total)
VBF	0.03±0.001 (0.27)	0.03±0.001 (0.20)	0.03±0.001 (0.20)	0.02±0.0009 (0.16)	0.01±0.0007 (0.12)	0.01±0.0007 (0.12)	0.01±0.0007 (0.12)	0.01±0.0007
GF	0.27±0.0049 (0.27)	0.07±0.003 (7.2×10^{-2})	0.07±0.002 (6.9×10^{-2})	0.03±0.002 (3.3×10^{-2})	0.009±0.0009 (9.4×10^{-3})	0.009±0.0009 (9.4×10^{-3})	0.009±0.0009 (9.4×10^{-3})	0.009±0.0009
$t\bar{t}$	1.28±0.09 (4.2×10^{-4})	1.06±0.08 (3.5×10^{-4})	1.05±0.08 (3.4×10^{-4})	0.47±0.05 (1.5×10^{-4})	0.08±0.02 (2.6×10^{-5})	0.08±0.02 (2.6×10^{-5})	0.08±0.02 (2.6×10^{-5})	0.08±0.02
WW	0.57±0.02 (3.5×10^{-3})	0.07±0.008 (4.4×10^{-4})	0.07±0.008 (4.2×10^{-4})	0.03±0.006 (2.1×10^{-4})	0.01±0.004 (6.8×10^{-5})	0.01±0.004 (6.8×10^{-5})	0.01±0.004 (6.8×10^{-5})	0.01±0.006
ZZ	0.71±0.01 (1.8×10^{-2})	0.11±0.006 (2.8×10^{-3})	0.11±0.006 (2.7×10^{-3})	0.04±0.005 (1.1×10^{-3})	0.007±0.0009 (1.8×10^{-4})	0.007±0.0009 (1.8×10^{-4})	0.007±0.0009 (1.8×10^{-4})	0.01±0.004
WZ	0.69±0.02 (5.1×10^{-3})	0.25±0.01 (1.8×10^{-3})	0.24±0.01 (1.8×10^{-3})	0.10±0.007 (7.4×10^{-4})	0.02±0.005 (1.4×10^{-4})	0.02±0.005 (1.4×10^{-4})	0.02±0.005 (1.4×10^{-4})	0.03±0.01
Z+jets	4.89±0.20 (7.1×10^{-5})	3.52±0.17 (5.1×10^{-5})	3.52±0.17 (5.1×10^{-5})	1.55±0.12 (2.2×10^{-5})	0.10±0.05 (1.5×10^{-6})	0.10±0.05 (1.5×10^{-6})	0.10±0.05 (1.5×10^{-6})	0.14±0.07
W+jets	1.63±1.30 (2.4×10^{-6})	0.31±0.52 (4.5×10^{-7})	0.15±0.37 (2.3×10^{-7})	0.15±0.37 (2.3×10^{-7})	-	-	-	3.83±2.24
Wbb+jets	-	-	-	-	-	-	-	0.35±0.27
Zbb+jets	0.52±0.04 (6.2×10^{-4})	0.40±0.03 (4.9×10^{-4})	0.40±0.03 (4.9×10^{-4})	0.17±0.02 (2.0×10^{-4})	0.01±0.01 (1.2×10^{-5})	0.01±0.01 (1.2×10^{-5})	0.01±0.01 (1.2×10^{-5})	0.02±0.02
QCD	-	-	-	-	-	-	-	505119±1078932
single top	-	-	-	-	-	-	-	0.07±0.09
Z→ ττ+jets	0.002±0.005 (5.5×10^{-8})	0.002±0.005 (5.5×10^{-8})	0.002±0.005 (5.5×10^{-8})	-	-	-	-	0.06±0.05
Total	10.28±1.32 (2.91×10^{-11})	5.73±0.56 (1.62×10^{-11})	5.55±0.42 (1.57×10^{-11})	2.51±0.39	0.23±0.06 (6.47×10^{-13})	0.23±0.06 (6.47×10^{-13})	0.23±0.06 (6.47×10^{-13})	0.29±0.07
Data	9	5	5	1	0	0	0	
$s/\sqrt{b}$	0.01±0.0004	0.011±0.0005	0.012±0.0004	0.013±0.0008	0.031±0.009	0.031±0.009	0.031±0.009	0.027±0.02

Table 6.4: Expected number of simulated events passing optimized high mass selection given an integrated luminosity of $33.4\,\mathrm{pb}^{-1}$, for the combined Z → $\mu\mu/ee$ channels. Efficiencies are given in brackets. The signal corresponds to the $\mathrm{m_H} = 300\,\mathrm{GeV}/c^2$ sample.

6.4.1 Mass dependence

A comparison between the baseline and optimized selections as a function of Higgs mass is shown in Figure 6.7. An improvement in significance (Z_1) of up to 10 % is observed as a function of Higgs mass. Using the optimized cuts, the low mass selection performs markedly better than the high mass selection for Higgs masses up to $m_H = 280\,GeV/c^2$. Beyond this the high mass selection provides better discrimination over background. This is largely due to increasing E_T^{miss} in higher Higgs mass signals. To maximise expected significance, one selection is adopted, with the low mass selection used for Higgs mass $m_H < 300\,GeV/c^2$ and the high mass selection for Higgs mass $m_H \geq 300\,GeV/c^2$. Adopting these selections, signal efficiency is maintained above 5% across all masses. However, it is seen to rise with increasing Higgs mass, to approximately 10 % in the low mass selection for $m_H = 280\,GeV/c^2$ and to around 20 % in the high mass selection for $m_H = 600\,GeV/c^2$. This is caused by the combination of increasing E_T^{miss} in the signal with increasing Higgs mass and the fixed E_T^{miss} cuts used. This is a favourable feature as the SM Higgs cross section decreases by over six orders of magnitude with increasing Higgs mass in the mass region considered. However, the change in cross section is the dominant effect, shown by the significance decreasing with increasing Higgs mass.

6.5 Conclusions

In this chapter an optimization of the baseline selection adopted in Chapter 5 was detailed. As in the baseline selection the main backgrounds resulting from the optimized selection are identified as Z+jets and $t\bar{t}$. The results of the optimization indicate an improvement in overall sensitivity of the analysis can be achieved in terms of signal significance, compared to the baseline result, of ≈10 % for the spectrum of masses considered between m_H = 200-600 GeV/c^2. However, it is likely this result will be affected by other uncertainties. Some important uncertainties are explored in the next chapter.

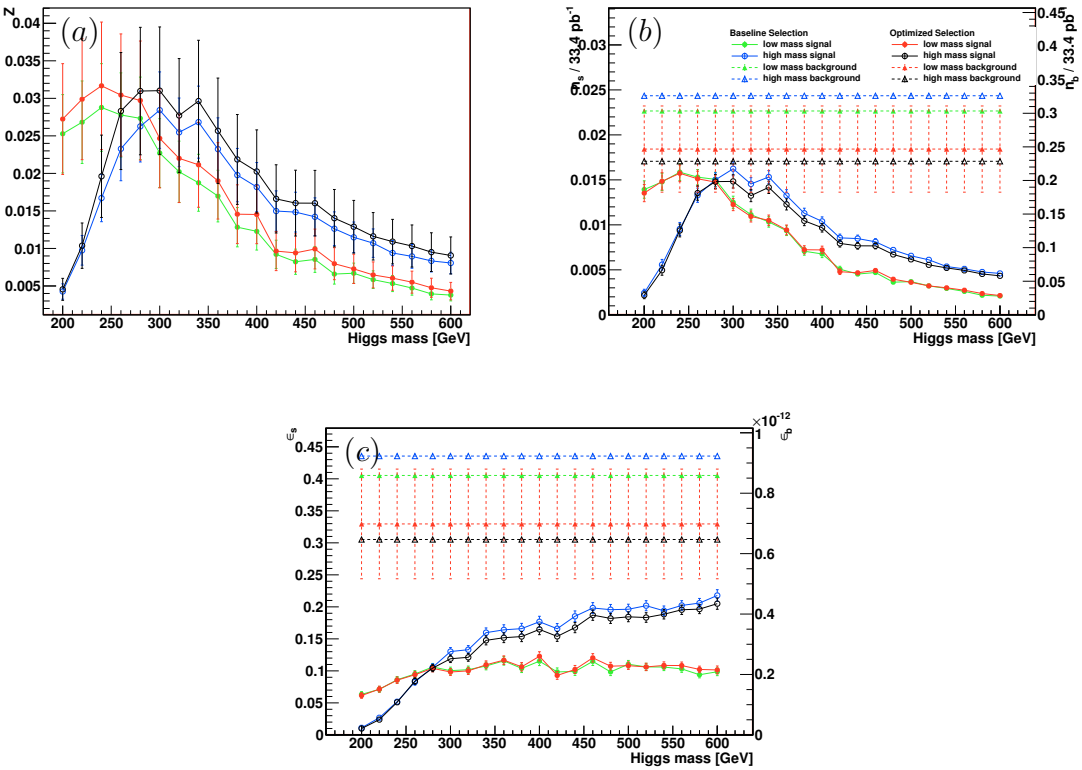


Figure 6.7: Comparison of results after applying baseline and optimized selections as a function of Higgs mass shown in terms of (a) significance (Z=Z_1),(b) expected number of signal (n_s) and background (n_b) events for an integrated luminosity of $33.4\,pb^{-1}$ and (c) signal (ϵ_s) and background (ϵ_b) efficiencies.

Chapter 7

Systematic Uncertainties and Final Sensitivity

This chapter explores a number of systematic effects which may lead to additional uncertainties on the search prescribed in the previous chapter. The systematic effects considered can be categorised into two types. The first type is those which give an uncertainty on the event yield and includes uncertainty on luminosity and uncertainty on signal and background cross sections. Methods to estimate the contribution of two of the main backgrounds, $t\bar{t}$ and Z+jets, are explored. The other type of uncertainty considered is that which results in an uncertainty on the acceptance efficiencies of the samples considered as a result of detector inefficiencies changing the kinematic distributions from which the analysis makes selections. In the studies presented these changes are explored in terms of their effect on the overall event yields. The combination of all systematic errors, added in quadrature, is used to define an overall systematic error in the rate of the total background and signal, for the low and high mass selections respectively. The chapter concludes with a discussion of the sensitivity of the search by way of 95 % C.L. limits on the signal cross section relative to the SM prediction as a function of Higgs boson mass for the 2010 dataset.

7.1 Theoretical uncertainties

The cross sections for the processes used in this analysis have associated uncertainties due to imperfect knowledge of the parton distribution functions, the incomplete description of the parton showering process and modelling of the underlying event and dependence on renormalisation and factorisation scales.

The uncertainty on the VBF process has been studied by the LHC Higgs cross section working group who estimate the uncertainty on the Standard Model Higgs boson cross section produced by VBF to be 3-9 % [19] in the mass range considered

Sample	Cross section uncertainty (%)	Effect on Total Background (%)	
		Low Mass Selection	High Mass Selection
signal	6	-	-
WW,WZ,ZZ	15	1.9	2.4
$t\bar{t}$	25	7.4	8.0
Z	5	2.5	4.8
Total Uncertainty	-	8.5	9.1

Table 7.1: Uncertainties on signal and background cross sections. Uncertainties on the signal rate were calculated using [19] and the background uncertainties taken from [41]. Z refers to Z and Zbb samples.

in this analysis. A flat uncertainty of 6 % is assumed in the following for each signal mass. The uncertainties on the background cross sections quoted in [41] are assumed. The uncertainties on the cross sections used are summarised in Table 7.1.

These uncertainties on the background cross sections, give rise to a total background uncertainty of ±8.5 % and ±9.1 % in the low and high mass selections respectively. As shown in Table 7.2, the increase in rate expected from considering di-boson+jets is within the uncertainty assumed on the nominal di-boson samples. The large decrease in rate observed for the WZ+jets samples is because only a subset of WZ decays are considered in the samples available.

Sample	Relative difference (%)	
	Low Mass Selection	High Mass Selection
WW	+14	-10
WZ	-65	-75
ZZ	+14	-5

Table 7.2: Comparison of yields of diboson and diboson+jets samples for optimized low and high mass selections. Relative difference corresponds to percentage difference in yield of diboson+jets samples compared to nominal diboson samples.

7.1.1 Data-driven background estimation

An important way in which the contribution from backgrounds in a search can be estimated, provided there is sufficient statistics, is by estimating them using data driven techniques. These techniques are merited as they allow reliance on predicted cross sections to be reduced and can also lead to reduced systematic uncertainties. Typically these methods rely on identifying control regions that are enriched with the backgrounds to be estimated by changing the selection cuts used in an analysis. In the following an attempt is made to use these techniques to estimate two of the main backgrounds in this analysis, $t\bar{t}$ and Z+jets, by reversing the cuts in the optimized selection, to identify $t\bar{t}$ and Z+jets control regions.

Ideally one would like to be able to estimate the contribution of the main backgrounds in an analysis in the region of cut space defined by the selection, i.e. in

the signal region. In order to estimate the contribution of a particular background in a signal region, a common approach is the ABCD method. In this method, the contribution of a background is estimated in the signal region A, by defining control (B and C) regions enriched with this background. A further control region, D corresponds to the control region after applying B and C changes. The number of background events in the signal region, N_A is then

$$N_A = \frac{N_B}{N_D} \times N_C \tag{7.1}$$

where N_B, N_C and N_D are the number of background events in the control regions B, C and D.

$t\bar{t}$ background

The method used to define the A,B,C and D regions for the estimation of the $t\bar{t}$ background is shown in Figure 7.1 The signal region is defined as that when the cuts

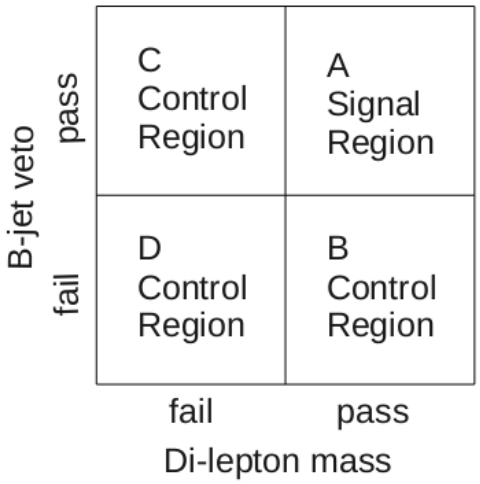


Figure 7.1: Definition of signal and control regions used to estimate the contribution of $t\bar{t}$ to the total background using di-lepton mass window and b-jet veto cuts.

relating to the Higgs decay products are made so as to maintain as much statistics as possible. In this signal region, $t\bar{t}$ remains a dominant background, contributing around 10 % to the total background as shown in the second column of Table 7.3.

As detailed in Chapter 5, it was shown that the b-jet veto was particularly efficient at rejecting $t\bar{t}$, reducing it by over 60% with respect to the previous cut applied. By explicitly requiring a b-tagged jet the control region $B^{t\bar{t}}$ enriched with $t\bar{t}$ may be established, as shown in the third column of Table 7.3 (again showing the numbers of expected events after the Higgs decay product selection). It shows

that although this control region consists mostly of $t\bar{t}$, approximately 70 %, there are other contributions to the total background, particularly from backgrounds with a Z. In order to try and reduce this background, one possibility is to reverse the di-lepton mass cut. To do this a second control region, $C^{t\bar{t}}$, is defined by reversing the di-lepton mass cut i.e. $|m_{ll} - m_Z| > 15\,\mathrm{GeV}/c^2$ but requiring $60 < m_{ll} < 150\,\mathrm{GeV}/c^2$. The result of this change in the selection, is shown in the fourth column of Table 7.3. In this region significant contamination from W+jets background becomes a problem. However, by requiring both a b-tagged jet and reversing the di-lepton mass window cut, a much purer (over 90 % of the total background) $t\bar{t}$ control region ($D^{t\bar{t}}$) is defined, as shown in the fifth column of Table 7.3. Using the number of $t\bar{t}$ expected in each control region, the contribution of $t\bar{t}$ in the signal region is predicted to be 3.94±0.19 compared to 3.69±0.15 measured, showing agreement within errors and indicating the methodology adopted works. If instead however, the total background values in each control region are used in the estimate of the $t\bar{t}$ contribution in the signal region, its value is overestimated. This is due to contributions from other backgrounds in the $B^{t\bar{t}}$ and $C^{t\bar{t}}$ control regions. For the same reason, using the ABCD method with the data leads to an over estimate of the predicted $t\bar{t}$ contribution in the signal region. However, if region $D^{t\bar{t}}$ is considered, a total background of 11.67±0.48 events are expected, of which 10.86±0.26 are $t\bar{t}$ and 10 events are observed in the data. This is within the 25 % normalisation error assumed for the $t\bar{t}$ background.

Sample	A^{tt}	B^{tt}	C^{tt}	D^{tt}
VBF	0.04±0.00	0.00±0.00	0.00±0.00	0.00±0.00
GF	0.26±0.01	0.00±0.00	0.01±0.00	0.00±0.00
$t\bar{t}$	3.69±0.15	6.96±0.21	6.14±0.19	10.86±0.26
WW	1.55±0.03	0.02±0.00	2.37±0.04	0.02±0.01
ZZ	1.41±0.02	0.04±0.01	0.03±0.00	0.00±0.00
WZ	1.59±0.03	0.03±0.01	0.16±0.01	0.00±0.00
Z+jets	16.82±0.37	0.89±0.09	1.81±0.13	0.04±0.04
W+jets	4.18±1.68	0.15±0.36	8.43±1.75	0.15±0.37
Wbb+jets	0.31±0.21	0.07±0.13	0.50±0.23	0.04±0.09
Zbb+jets	1.44±0.06	1.39±0.06	0.13±0.02	0.08±0.02
QCD	2.12±2.88	0.00±0.00	0.00±0.00	0.00±0.00
single top	0.11±0.10	0.21±0.11	0.39±0.11	0.46±0.12
Z→ $\tau\tau$+jets	0.21±0.05	0.00±0.00	0.37±0.07	0.01±0.02
Total	33.45±3.37	9.76±0.46	20.33±1.78	11.67±0.48
Data	36.00	9.00	18.00	10.00

Table 7.3: Number of normalised events ($33.4\,\mathrm{pb}^{-1}$) after the full selection for the $m_H = 200\,\mathrm{GeV}/c^2$ signal, each background and total background. Included is the observed data events. The number of events are quoted for the nominal selection $A^{t\bar{t}}$ and top control regions B,C,$D^{t\bar{t}}$.

Z+jets background

In the following the ABCD method is used to estimate the contribution of Z+jets background (assuming negligible contribution from Z→ $\tau\tau$ as observed). The two

most discriminating cuts against this background were identified as the $E_{\mathrm{T}}^{\mathrm{miss}}$ cut and the $\Delta\eta$ cut. To this end, $\mathrm{B^Z}$ is defined as the control region where $E_{\mathrm{T}}^{\mathrm{miss}} < 40\,\mathrm{GeV}$ is required and $\mathrm{C^Z}$ that in which $\Delta\eta < 3.6$. Application of both of these changes to the selection gives control region $\mathrm{D^Z}$. The signal region $\mathrm{A^Z}$, is defined as that at the end of the full (low mass) selection, because of the use of $\Delta\eta$ cut to define the $\mathrm{C^Z}$ control region. Because data and non pile-up Monte Carlo were shown not to be in agreement for $E_{\mathrm{T}}^{\mathrm{miss}} < 40\,\mathrm{GeV}$, in the following no comparison with data is made. Instead, the suitability of the control regions identified in estimating the Z+jets contribution to the signal region is explored using Monte Carlo only. The results of defining the Z+jets control regions in this way is presented in Table 7.4 in terms of the number of normalised events surviving the full selection in each case. 0.51±0.02 Z events in the signal region are predicted using the number of Z+jets events in the $\mathrm{B^Z}$,$\mathrm{C^Z}$ and $\mathrm{D^Z}$ control regions. This overestimates the contribution observed and is a consequence of the ratio $\mathrm{A^Z/C^Z}$ not being close to $\mathrm{B^Z/D^Z}$ as required for the ABCD method to work. This is because the cuts defining the control regions are not completely independent. Both the $\mathrm{B^Z}$ and $\mathrm{D^Z}$ control regions provide a high purity Z+jets sample (over 90 % total background) whereas the $\mathrm{C^Z}$ control region suffers from large background contributions, particularly $\mathrm{t\bar{t}}$.

Sample	$\mathrm{A^Z}$	$\mathrm{B^Z}$	$\mathrm{C^Z}$	$\mathrm{D^Z}$
VBF	0.02±0.00	0.01±0.00	0.02±0.00	0.01±0.00
GF	0.01±0.00	0.01±0.00	0.06±0.00	0.04±0.00
$\mathrm{t\bar{t}}$	0.21±0.04	0.05±0.02	2.89±0.13	1.30±0.09
WW	0.03±0.01	0.02±0.00	0.18±0.01	0.11±0.01
ZZ	0.02±0.00	0.13±0.01	0.20±0.01	5.16±0.09
WZ	0.04±0.01	0.31±0.02	0.49±0.02	7.79±0.09
Z+jets	0.30±0.08	71.20±0.76	8.12±0.26	1143±3
W+jets	0.00±0.00	0.00±0.00	0.46±0.56	1.23±0.84
Wbb+jets	0.00±0.00	0.00±0.00	0.11±0.13	0.16±0.18
Zbb+jets	0.03±0.02	3.02±0.08	0.93±0.05	40.53±0.30
QCD	0.00±0.00	1.06±2.52	0.00±0.00	5.47±3.69
single top	0.00±0.00	0.01±0.04	0.06±0.07	0.08±0.08
$\mathrm{Z}\rightarrow\tau\tau$+jets	0.00±0.00	0.01±0.02	0.06±0.04	0.20±0.05
Total	0.62±0.09	75.82±2.63	13.51±0.65	1205±5

Table 7.4: Number of normalised events ($33.4\,\mathrm{pb}^{-1}$) after the full selection for the $\mathrm{m_H} = 200\,\mathrm{GeV}/c^2$ signal, each background and total background. The number of events are quoted for the nominal selection $\mathrm{A^Z}$ and Z+jets control regions B,C,$\mathrm{D^Z}$.

7.2 Experimental uncertainties

In addition to the theoretical uncertainties described previously, it is expected that uncertainties relating to detector performance, mainly caused by misalignment, extra material in the detector and mis-calibration, will be relevant for this analysis as it relies heavily on all the major sub-detectors of the detector. In the following

several sources of uncertainty are investigated. The uncertainties used are derived from the relevant performance group or where not available follow [41]. Where possible, official tools provided by the performance groups have been used to assess their effect. In each case the effect of the uncertainty is measured by repeating the analysis for each systematic effect in turn and comparing the overall percentage difference in yield for each background and the total background. The uncertainty on the signal is estimated by repeating the analysis for the $m_H = 200{,}400{,}600\,\mathrm{GeV}/c^2$ signal samples. The final signal uncertainties used come from the $m_H = 200\,\mathrm{GeV}/c^2$ sample for the low mass selection and the average of the results from the $m_H = 400{,}600\,\mathrm{GeV}/c^2$ samples for the high mass selection.

7.2.1 Luminosity

The uncertainty on the luminosity has been measured in [85] and is estimated to be 3.4 %.

7.2.2 Energy scale

The energy scale of electrons and muons is expected to affect the selection relating to the Higgs decay products. In particular it has been shown that the di-lepton mass window cut provides lots of discrimination over background, which may be worsened when uncertainty on electron and muon energy scales is taken into account. However, it is expected that this effect will be small because of the high accuracy with which electrons and muons are measured. To estimate the effect of the uncertainty on the electron and muon energy scales, the analysis was repeated and the energy scale of electrons and muons systematically shifted up and down independently. For electrons, as recommended by the eγ performance group [64], a 1(3) % uncertainty on the energy scale of electrons in the barrel(end-cap) was assumed. For muons the energy scale uncertainty was assumed to be 1 %.

In the same way that the energy scale of electrons and muons may affect the acceptance of the analysis with regard to variables related to the Higgs decay products, similarly uncertainty on the jet energy scale is likely to affect the selection related to the VBF remnants. For this reason, the analysis was repeated varying the jet energy scale up and down according to the uncertainty recommendations from the jetE_T^{miss} performance group. This is implemented using the official Jet Energy Uncertainty Provider [86, 87] which provides a correction to the energy scale of a jet as a function of its p_T and η. Typically, the jet energy uncertainty for jets with $p_T < 100\,\mathrm{GeV}/c$ is 8(9)% and 6(7)% for jets with $p_T > 100\,\mathrm{GeV}/c$ in the central(endcap) region [86].

When changing the energy scale of the objects discussed, the effect was propagated to $E_{\mathrm{T}}^{\mathrm{miss}}$ by recalculating the $E_{\mathrm{T}}^{\mathrm{miss}}$ components after modifying the reconstruction performance of the relevant physics object. The effect of changing the electron, muon and jet energy scales is summarised in Table 7.5. The effect of changing the jet energy scales has the largest effect. This is largely because the uncertainty on the jet energy scale is greatest. For the same reason the effect of changing the electron energy scale is larger than changing the muon energy scale. In the signal it is clear

	Low Mass Selection						High Mass Selection					
	Electrons		Muons		Jets		Electrons		Muons		Jets	
Sample	Up	Down	Up	Down	Up	Down	Up	Down	Up	Down	Up	Down
$m_H = 200$	0.98	-1.46	-0.51	0.00	6.96	-16.76	-2.92	-3.02	0.00	-6.21	24.30	-18.20
$m_H = 400$	-0.31	-1.29	0.34	0.00	2.24	-12.47	0.01	-0.72	0.50	-0.25	-0.46	-8.46
$m_H = 600$	0.01	-0.36	-0.37	-0.37	4.01	-4.35	0.01	-0.54	0.18	-0.18	0.73	-5.71
top	6.99	-7.13	0.00	0.00	10.32	-13.68	-0.03	15.02	7.79	0.00	-0.40	-15.37
WW	0.01	-9.56	0.00	0.00	38.74	-29.05	6.16	-0.01	0.00	-6.21	-6.28	-18.80
ZZ	0.01	-1.86	0.00	0.00	29.90	-28.55	0.01	-1.46	0.00	0.00	10.85	-17.44
WZ	2.63	-0.01	0.76	0.36	15.67	6.03	0.01	-8.15	0.00	-1.55	-7.18	-11.55
W+jets	-	-	-	-	-	-	-	-	-	-	-	-
Z+jets	12.84	-11.61	13.34	-6.56	-12.45	55.71	0.00	-0.00	7.31	0.00	-15.81	44.44
Wbb+jets	-	-	-	-	-	-	-	-	-	-	-	-
Zbb+jets	0.01	-0.01	20.01	0.00	39.82	-16.32	0.01	-0.01	16.51	0.00	-15.69	-2.99
QCD	-	-	-	-	-	-	-	-	-	-	-	-
Total	8.98	-8.55	7.57	-3.27	1.72	21.63	0.29	4.43	6.67	-0.43	-8.48	12.03

Table 7.5: Relative changes (%) in efficiency of the optimized selections for m_H = 200, 400, 600 GeV/c^2 signal samples, each individual background and total background when the energy scale of electrons, muons and jets is varied as detailed in Section 7.2.2. The table includes values for the optimized low and high mass selections. Values are in %.

that the change in event yield is dominated by the effect of the jet energy scale, in each case the effect of the electron and muon energy scale uncertainties is always less than 1%. To this end, for the signal the effect of the electron and muon energy scales is neglected. In addition, because the effect of jet energy scale uncertainty is by far the dominant effect compared to all other detector related uncertainties investigated for the signal, these other effects are neglected. For the low mass selection, increasing the jet energy scale gives rise to an increase in efficiency of around 7%, whereas decreasing the jet energy scale sees a decrease in events by around 17 %. A value of ± 12 % is assumed. For the high mass selection, the signal appears to be less affected by changes in the jet energy scale. An uncertainty of ± 5.5 % on the signal rate due to jet energy scale uncertainty in the high mass selection is used.

For the background, the effect of the electron/muon energy scale uncertainty is seen to have a non-negligible effect on total background rate. In the low mass selection, for electrons, a roughly symmetric change in total background yield, of ± 9 % is observed when increasing/decreasing the electron energy scale. The same trend is true when changing muon energy scale, although the shift down in efficiency

is approximately half the effect of the shift up. A value in between these two results is assumed and as such the background uncertainty due to muon energy scale uncertainty is approximated as ±5 %. Considering the events selected in the analysis either contain two electrons or muons it does not make sense to consider the electron and muon energy scale uncertainties independently. As such, a conservative estimate for the total background uncertainty of ±9 % is assumed in the low mass selection, due to the combined electron and muon energy scale uncertainties.

In the high mass selection, the effect of changing the electron and muon energy scales is less pronounced. Following a similar argument to that used for the low mass selection, the combined electron and muon energy scale uncertainty gives an uncertainty on the total background rate in the high mass selection of ±3.5 %.

The effect of changing jet energy scale appears to have a large effect on the total background rate, giving rise to an increase in total background of over 20 % in the low mass selection in the case of decreasing jet energy scale. The effect of increasing background rate with decreasing jet energy scale is caused by the CJV cut accepting more events in this scenario, as illustrated in Figure 7.2. The largest effect of this kind is seen for the Z+jets background, which increases by over 50 %. Since this is a major background, this dominates the overall change in background, giving rise to the net increase in total background rate observed. In contrast, when increasing the jet energy scale a much lower change on total background is observed, of a few percent. In this case, despite all other background rates increasing, a net decrease in Z+jets background is observed. This is again caused by the CJV cut, in this case leading to an increased suppression of high jet multiplicity Z events. Again because it is a dominant background, the decrease in Z rate compensates for the increase in background rate from other processes, giving a smaller net change in total background.

In the high mass selection a similar trend is observed, although the increase in background rate by decreasing jet energy scale is less pronounced, guided by a lower increase in Z+jets background. The effect of increasing jet energy scale leading to a net decrease in background rate is now observed for both $t\bar{t}$ and Z, driving the observed net decrease in total background of 9 %. Applying an approximate average to the shifts observed, the uncertainty on the total background rate due to jet energy scale uncertainty, for the low and high mass selections is taken to be ±10 %.

7.2.3 Energy resolution

In the following the effect of electron, muon and jet energy resolution uncertainties on the analysis are investigated. In each case, re-calculation of $E_{\mathrm{T}}^{\mathrm{miss}}$ components was made in the same way as was done when looking at the effect of energy scale. In

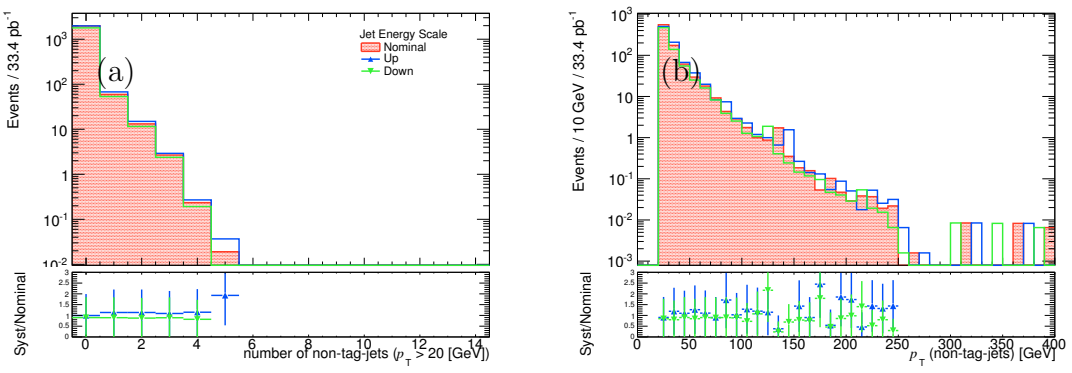


Figure 7.2: Distributions of (a) non tag-jet multiplicity and (b) non tag-jet p_{T} for the total background (after requirement of two leptons), showing the effect increasing(Up) and decreasing(Down) jet energy scale compared to nominal case. The ratio shows this effect in terms of systematic change (Syst) divided by nominal case (Nominal).

order to estimate the effect of electron energy resolution, the recommendations from the $e\gamma$ performance group were used [64]. The energy of electrons was corrected by adding ΔE, found using a randomly distributed Gaussian with width σ

$$\sigma = \sqrt{(S(1+S_{error})\sqrt{E_{cluster}})^2 + (C(1+C_{error})E_{cluster})^2 - (S\sqrt{E_{cluster}})^2 - (CE_{cluster})^2} \tag{7.2}$$

where S=0.2 represents the electromagnetic calorimeter sampling term, C=0.007 its constant term and S_{error}=±20 % and C_{error}=±100(400) %($|\eta_{cluster}| < (\geq)$ 1.37) their respective errors. $E_{cluster}$ is the cluster energy.

The effect of changing the p_{T} resolution of muons was investigated by smearing the p_{T} of muons measured in the inner detector and muon spectrometer. This was done using the same method as described in Section 4.3.1, but instead retrieving the result after applying a $+1\,\sigma$ variation. The effect of varying the jet energy resolution was investigated using the recommendations from the Jet$E_{\mathrm{T}}^{\mathrm{miss}}$ group. This provides a correction to a jet's energy based on its p_{T} and η.

The effect of the uncertainty on the electron and muon energy resolutions is of similar magnitude in the low mass selection. A value of +5 % uncertainty on total background rate is assumed for the combined electron and muon energy resolution uncertainty. In contrast the uncertainty on jet energy resolution leads to a -9 % shift in total background. Because the selection used in this study requires electrons/muons and jets and the corresponding uncertainties stated above are one sided and have opposite effects, the overall effect due to the combined energy resolution uncertainty of electrons/muons and jets is a -4 % uncertainty on total background

rate. The associated systematic uncertainty is therefore taken as ±2 %. Using a similar approach for the high mass selection an effect of -2 % is observed, taking into account the +5 % increase and -7 % decrease in total background expected from the uncertainty due the electron and muon/jet energy resolution respectively. The associated systematic uncertainty is taken to be ±1 %.

	Low Mass Selection			High Mass Selection		
Sample	Electrons	Muons	Jets	Electrons	Muons	Jets
$m_H = 200$	-0.97	1.02	-1.49	-0.01	0.00	-0.16
$m_H = 400$	-0.65	0.34	-1.94	-0.48	0.25	-1.70
$m_H = 600$	0.01	0.00	-1.45	0.00	0.18	-0.70
top	-0.07	0.00	-21.45	14.75	0.00	-7.79
WW	-9.55	0.00	0.03	6.15	0.00	0.00
ZZ	-0.00	-0.30	-4.63	-1.45	0.00	-3.56
WZ	-0.36	0.74	1.61	-1.84	-0.01	3.02
W+jets	-	-	-	-	-	-
Z+jets	7.51	13.16	-5.23	-0.00	0.00	-7.91
Wbb+jets	-	-	-	-	-	-
Zbb+jets	15.73	0.00	22.95	-0.00	0.00	-15.12
QCD	-	-	-	-	-	-
Total	4.08	6.65	-8.91	5.17	-0.00	-6.73

Table 7.6: Relative changes in efficiency for $m_H = 200, 400, 600\,\mathrm{GeV}/c^2$ signal samples, each individual background and total background when the energy resolution of electrons, muons and jets is varied as detailed in Section 7.2.3. The table includes values for the optimized low and high mass selections. Values are in %.

7.2.4 Reconstruction efficiency

It was noted in Chapter 4 that the Monte Carlo reconstruction efficiency is overestimated compared to that in the data and as such a correction was applied. No correction was applied for muon efficiency. The effect of the uncertainty on the reconstruction and identification efficiency of electrons and muons on the analysis was investigated assuming a uncertainty of 2.3(1) % respectively. This was implemented by applying an additional 4.6(2) % additional weighting up and down for events satisfying the lepton selection in the $Z \to ee(Z \to \mu\mu)$ channels respectively. These uncertainties were taken from [41].

The effect of changing the efficiencies is shown in Table 7.7. Overall the effect on the total background rate is of the order a few percent, for both the low and high mass selections. The combined uncertainty on total background rate due to electron and muon reconstruction efficiency uncertainty is taken to be ±2 %.

7.2.5 b-tagging efficiency

The uncertainty on b-tagging efficiency is expected to be important for this analysis as one of the main backgrounds is top, involving decays to b hadrons. Efficient

	Low Mass Selection				High Mass Selection			
Sample	Electrons		Muons		Electrons		Muons	
	Up	Down	Up	Down	Up	Down	Up	Down
$m_H = 200$	1.53	-1.55	1.42	-1.44	1.28	-1.30	1.67	-1.70
$m_H = 400$	1.38	-1.40	1.22	-1.25	1.40	-1.42	1.21	-1.23
$m_H = 600$	1.17	-1.19	1.35	-1.38	1.28	-1.30	1.24	-1.27
top	2.85	-2.86	1.63	-1.63	1.32	-1.32	3.16	-3.16
WW	1.32	-1.35	1.93	-1.97	1.81	-1.84	1.16	-1.21
ZZ	0.42	-0.42	2.00	-2.06	0.87	-0.88	1.57	-1.61
WZ	1.46	-1.48	1.28	-1.31	1.76	-1.79	0.94	-0.96
W+jets	-	-	-	-	-	-	-	-
Z+jets	0.95	-0.98	1.77	-1.85	0.46	-0.48	2.16	-2.25
Wbb+jets	-	-	-	-	-	-	-	-
Zbb+jets	1.57	-1.61	1.26	-1.31	1.09	-1.12	1.66	-1.72
QCD	-	-	-	-	-	-	-	-
Total	1.65	-1.67	1.68	-1.72	0.97	-0.99	2.31	-2.36

Table 7.7: Relative changes in efficiency for $m_H = 200$, 400, 600 GeV/c^2 signal samples, each individual background and total background when the combined reconstruction and identification efficiency of electrons and muons is varied as detailed in Section 7.2.4. The table includes values for the optimized low and high mass selections. Values are in %.

rejection of this background relies on accurate b-tagging algorithms. The uncertainty on the b-tagging efficiency is investigated using the recommendations from the b-tagging performance group [88]. In this method each jet is assigned a (p_T and η dependent) scale factor. The scale factor takes on a different form depending on whether the jet is tagged by the SV0 algorithm (i.e. has SV0 weight > 5.72 in this analysis) or not. In addition the form of the scale factor is different depending on the Monte Carlo truth nature of the associated jet, i.e. if it is identified as derived from b quarks. In this analysis the procedure of identifying the truth origin of each jet was estimated using the quark flavour of the closest matched truth quark in ΔR.

The effect of the b-tagging efficiency uncertainty on the distribution of the highest jet SV0 weight per event for the total background is shown in Figure 7.3. As expected, increasing the b-tagging efficiency leads to the weight of jets being made larger, indicating an expected decrease in total background because of the b-jet veto cut. The opposite scenario is shown to occur when decreasing the b-tagging efficiency.

These results are expressed in terms of overall relative shift on the signal and background rate in Table 7.8. Changing the b-tagging efficiency is shown to have negligible effect on all processes considered apart from top and Zbb backgrounds, for which increases of up to 30,10 % in their respective yields is expected by decreasing the b-tagging efficiency. The overall change in the total background rate is expected to be more pronounced in the high mass selection by a few percent to approximately 10 %. This is explained by the contribution of top to the total background in the high mass selection being larger compared to that in the low mass selection. Taking into

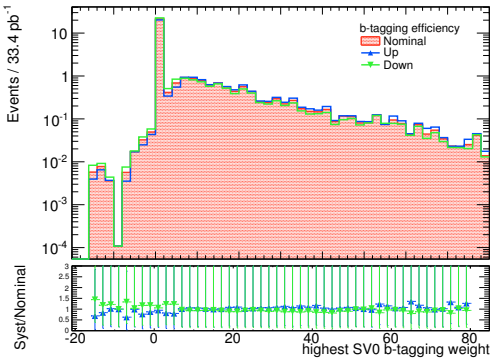


Figure 7.3: Effect of increasing (Up) and decreasing (Down) b-tagging efficiency on the distribution of highest SV0 b-tag weight for the total background, compared to the nominal case (Nominal). The ratio shows this effect in terms of systematic change (Syst) divided by nominal case (Nominal).

account the differences when shifting up and down, the uncertainties on the total background rate for the low and high mass selections, due to b-tagging efficiency uncertainty, are taken to be ±7.5 % and ±9 % respectively.

	Low Mass Selection		High Mass Selection	
Sample	Up	Down	Up	Down
$m_H = 200$	-0.77	0.38	-0.44	-0.16
$m_H = 400$	-1.04	0.19	-0.81	0.02
$m_H = 600$	-0.56	0.23	-0.73	0.40
top	-13.44	22.33	-20.50	29.27
WW	-3.40	2.04	-1.98	1.18
ZZ	-1.92	1.47	-1.51	0.81
WZ	-1.29	0.71	-1.89	1.06
W+jets	-	-	-	-
Z+jets	-1.16	1.12	-0.55	0.05
Wbb+jets	-	-	-	-
Zbb+jets	-13.11	13.11	-15.12	8.25
QCD	-	-	-	-
Total	-5.88	8.77	-8.26	10.62

Table 7.8: Relative changes in efficiency for $m_H = 200,400,600\,\mathrm{GeV}/c^2$ signal samples, each individual background and total background when the b-tagging efficiency is varied as detailed in Section 7.2.5. The table includes values for the optimized low and high mass selections. Values are in %.

7.2.6 Total systematic uncertainty

The sources of systematic uncertainty investigated and their associated values for the signal and total background, for the low and high mass selections are summarised in Table 7.9. The total systematic uncertainty on the signal and total background rate in each selection is obtained by adding the rate uncertainties derived from considering each systematic uncertainty in quadrature. In the low mass selection, the total estimated systematic uncertainty on the signal rate is ±14 % and on the total

Uncertainty	Signal		Background	
	Low mass	High Mass	Low Mass	High Mass
Luminosity	±3.4	±3.4	±3.4	±3.4
Cross Sections	±6	±6	±8.5	±9.1
Electron and Muon Efficiency	-	-	±2	±2
Electron, Muon and Jet Energy Resolution	-	-	±2	±1
Electron Energy and Muon Energy scale	-	-	±9	±3.5
Jet Energy scale	±12	±5.5	±10	±10
b-tagging	-	-	±7.5	±9
Total	±14	±9	±18	±17

Table 7.9: Summary of each systematic effect considered on the signal and background for the low and high mass selections. Values are in %.

background rate it is ±18 %. For the high mass selection the estimated systematic uncertainty on the signal rate is ±9 % and on the total background rate is ±17 %. The effect of these uncertainties on the signal significance Z=s/$\sqrt{b}$ is shown in Figure 7.4. The largest impact is observed where the low mass selection is applied, for Higgs masses less than 300 GeV/c^2. Here approximately a 5 % change in signal significance is observed. A negligible impact is observed for masses above 300 GeV/c^2, owing to more similar uncertainties on signal and background rates for the high mass selection. With improvements in understanding of the detector response and modelling of physics interactions, these uncertainties will be reduced.

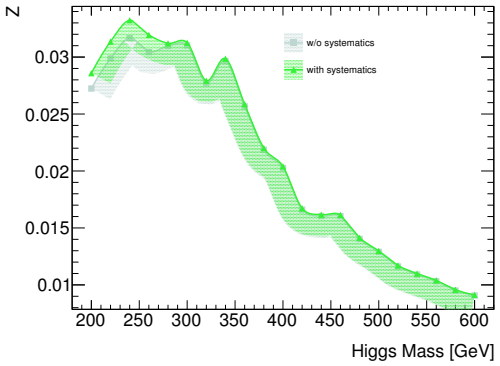


Figure 7.4: Effect of systematics on signal significance as a function of Higgs mass [GeV].

7.2.7 Comment on effect of pile-up

The selection with regard to variables relating to the Higgs decay products has been shown in Section 6.4 to be largely independent of the pile-up levels seen in the 2010 data, for the selection defined. However, discrepancies were observed when considering the VBF selection and this was attributed to pile-up because of the better observed agreement between data and pile-up reweighted Monte Carlo for the associated variable distributions. Because of insufficient statistics in the pile-up samples available, a detailed study of the effect of pile-up cannot be made with the full selection. In the following an attempt is made to estimate the effect of pile-up on the analysis by comparing the final number of events after applying the (low mass) selection relating to the VBF selection only. The results of this are presented in Table 7.10. It indicates that pile-up at the level considered gives rise to +21 % shift in total background rate, largely due to a large net increase in Z+jets background. This highlights the importance that pile-up is likely to have in this channel and that further studies will need to be done to better estimate its effect. With regard to this analysis, it was observed that the main discrepancy between non pile-up and pile-up Monte Carlo was due to the $\Delta\eta$ distribution. Since this is included in the high mass selection (where no CJV or cut on tag-jets mass is made) the +21 % shift in total background rate is assumed for both the low and high mass selections. This is applied not as a systematic but as a shift in the total background rate. No change in the signal rate for the masses tested was observed. The effect on the signal significance Z=s/$\sqrt{b}$ is shown in Figure 7.5. As expected the effect is to reduce signal significance. For Higgs masses lower than 300 GeV/c^2, where the low mass selection is applied a larger effect is observed because of the slightly higher background level.

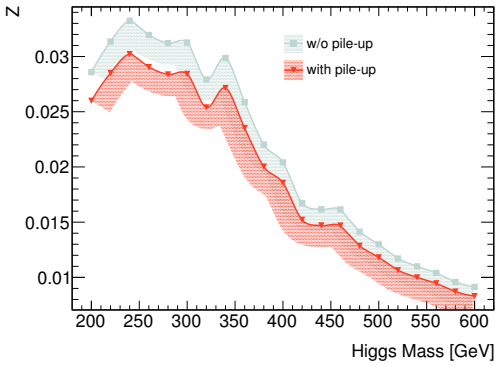


Figure 7.5: Estimated effect of pile-up on signal significance as a function of Higgs mass [GeV].

Sample	Non pile-up	pile-up	Relative change
$m_H = 200$	0.03±0.001(0.11)	0.03±0.001 (0.12)	0
$m_H = 400$	0.01±0.0005 (0.18)	0.01±0.0005 (0.18)	0
$m_H = 600$	0.01±0.0005 (0.18)	0.01±0.0005 (0.18)	0
top	3.32±0.17 (9.5×10^{-4})	3.40±0.15 (9.7×10^{-4})	2.4
WW	0.08±0.0009 (5.0×10^{-4})	0.08±0.0008 (5.0×10^{-4})	-
ZZ	0.09±0.01 (2.3×10^{-3})	0.09±0.01 (2.2×10^{-3})	-
WZ	0.31±0.02 (2.3×10^{-3})	0.32±0.02 (2.4×10^{-3})	3
Z+jets	32.21±0.51 (3.1×10^{-4})	44.13±2.88 (4.3×10^{-4})	27
W+jets	0.46±0.64 (6.8×10^{-7})	0.50±0.57 (7.3×10^{-7})	8
Wbb/c+jets	-	-	-
Zbb+jets	2.20±0.07 (1.8×10^{-3})	1.98±0.08 (1.6×10^{-3})	10
QCD	6.77±3.92 (1.9×10^{-11})	6.77±3.92 (1.9×10^{-11})	0
Total	45.4±4.01 (1.3×10^{-10})	57.3±4.42 (1.6×10^{-10})	20.8

Table 7.10: Normalised number of events (efficiency in brackets) and corresponding relative changes in efficiency for $m_H = 200{,}400{,}600\,\mathrm{GeV}/c^2$ signal samples, each individual background and total background, when the VBF related selection is compared in pile-up reweighted Monte Carlo (apart from QCD where non pile-up samples were used) and non pile-up Monte Carlo. Values correspond to the optimized low mass selection. Relative changes are in %.

7.3 Limits

Typically the results of a search are analysed by making a statement about how well the observed results agree or disagree with different hypotheses. In a search for a signal, the null (background only) hypothesis corresponds to the absence of this signal and the alternative corresponds to the signal plus background hypothesis. A test statistic is used to distinguish between the different hypotheses. In the following this is based on the profile likelihood ratio. The total number of observed events (signal s, background b) is assumed to be distributed according to a Poisson distribution with mean μs+b, where $\mu = \sigma/\sigma_{SM}$ is the signal strength with $\mu = 0$ corresponding to the absence of signal and $\mu = 1$ to a signal rate as expected by the SM. Free parameters other than μ are referred to as nuisance parameters. Because of the low number of expected signal and background events, in this study the only nuisance parameter (θ) considered is the expected number of background events. It is also assumed to be distributed according to a Poisson distribution. A likelihood function $L(\mu, \theta)$ is then defined as the product of these Poisson probabilities and the profile likelihood ratio as

$$\lambda(\mu) = \frac{L(\mu,\hat{\hat{\theta}})}{L(\hat{\mu},\hat{\theta})} \qquad (\mu \geq 0) \tag{7.3}$$

where $\hat{\mu}$ and $\hat{\theta}$ represent the maximum likelihood estimators of μ and θ. $\hat{\hat{\theta}}$ denotes the conditional maximum likelihood estimator of θ when maximizing L for a specific

value of $\hat{\mu}$. The test statistic is then defined by

$$q_\mu = -2 \ln \lambda(\mu) \tag{7.4}$$

The upper limit on the signal cross section is calculated for each tested Higgs mass point using the CL_s method [89], in which the CL_s value is defined as

$$CL_\mathrm{s} = \frac{CL_\mathrm{s+b}}{CL_\mathrm{b}} \tag{7.5}$$

where $CL_{s+b}(\mathrm{CL}_b)$ is the probability under the assumption of the signal plus background (background) hypothesis that a value of the corresponding test statistic is found with equal or lesser compatibility with the signal plus background (background) model compared to the value of the test statistic observed. This corresponds to the p-value of the signal plus background (background) hypothesis. The upper limits are set to the 95 % C.L. on the signal cross section at each Higgs mass by evaluating the CL_s value for a range of μ values and finding the corresponding μ such that the CL_s value converges to 0.05. If this μ has a value of less than one the signal is regarded as excluded at the 95 % C.L.

The 95 % C.L. limits on the signal cross section relative to the SM prediction are shown in Figure 7.6 with the values tabulated in Table 7.11. The integrated luminosity corresponds to $33.4\,\mathrm{pb}^{-1}$ of data collected in 2010. The most stringent limits are found in the mass region between 250 and $350\,\mathrm{GeV}/c^2$. This corresponds to the high mass end of the low mass selection, giving the best compromise between increased signal discrimination and the reduction in signal cross section with increasing Higgs mass. Outside this region, with decreasing Higgs mass the limits are worsened by the poorer signal over background discrimination of the cuts used while for higher masses, the same trend is observed but is largely due to decreasing signal cross section with increasing Higgs mass. Overall the limits are slightly better than 1/10 the expected confidence limits obtained for the gluon fusion analysis [41]. This indicates that the VBF specific analysis considered here is at least as competitive as the GF analysis, when the order of magnitude lower cross section for the VBF signal is taken into account.

m_H	Observed	Expected	-1σ	+1σ	+2σ
200	252.6	254.1	2.2	86.3	230.8
220	224.0	229.2	3.7	86.4	211.9
240	215.3	218.1	3.0	78.9	201.9
260	228.5	229.7	1.3	82.3	205.9
280	224.6	227.0	3.2	74.9	173.3
300	226.3	228.5	2.8	72.2	174.0
320	253.1	258.7	7.7	78.4	183.0
340	233.0	235.8	4.4	83.5	185.3
360	271.9	276.8	4.6	89.6	199.7
380	315.5	317.3	4.5	111.0	243.6
400	344.0	348.1	8.2	114.1	252.7
420	408.8	413.4	6.2	138.3	301.4
440	429.7	432.6	3.8	126.1	299.5
460	433.0	434.9	4.3	137.3	294.8
480	481.7	488.1	6.5	143.1	298.6
500	515.2	517.1	3.5	155.6	303.1
520	561.8	570.5	5.4	162.9	328.4
540	601.5	603.0	6.5	170.2	397.0
560	630.1	630.4	3.6	186.3	369.6
580	673.7	677.8	8.4	203.7	322.2
600	720.5	730.3	11.2	187.6	269.7

Table 7.11: Expected and observed exclusion limits (95% C.L.) on the ratio of the VBF signal cross section to the SM cross section for the range of Higgs boson masses [GeV] considered in this analysis, including $\pm 1\sigma$,$+2\sigma$ errors on the expected values. The integrated luminosity corresponds to $33.4\,\mathrm{pb}^{-1}$ of data collected in 2010.

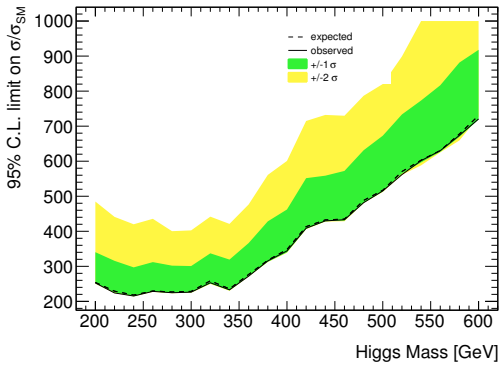


Figure 7.6: Expected and observed VBF signal cross section exclusion limits relative to the predicted SM cross section as a function of Higgs boson mass [GeV] taking into account the optimized selection with the effect of pile-up (systematic uncertainties quoted in Section 7.2.6 also taken into account). Integrated luminosity corresponds to $33.4\,\mathrm{pb}^{-1}$ of data collected in 2010. The dashed line shows the expected exclusion while the green and yellow bands show the associated $\pm 1\sigma$ and $\pm 2\sigma$ statistical errors. The solid line represents the observed limits.

7.4 Conclusions

In this chapter a number of systematic uncertainties have been investigated and their impact on the analysis developed studied. The largest impacting individual systematic uncertainty was found to be jet energy scale uncertainty, giving rise to a total background uncertainty of the order of 10 %. However it was also shown that cross section uncertainty has a large effect. Two methods were explored to try and estimate the background contribution from the main backgrounds identified. Each method explored relies on identifying control regions enriched with the background to be estimated. The method used to estimate $t\bar{t}$ showed by using control regions made from reversing the b-tagging and di-lepton mass window cuts, a very pure $t\bar{t}$ control sample could be produced. Although this could not be confirmed in the signal region, using this control region directly, the uncertainty assumed on the total $t\bar{t}$ cross section was confirmed as suitable. The method to estimate Z+jets used control regions by reversing the $E_{\mathrm{T}}^{\mathrm{miss}}$ and $\Delta\eta$ cuts. Although this was shown to produce a very pure Z+jets sample, the non independent nature of the cuts used to identify the control regions meant that this method did not work. The largest individual uncertainty on this analysis is that due to pile-up. However, the suggested estimate of its effect is approximate and further studies using pile-up Monte Carlo are needed to understand this better. Comparing systematic uncertainty for the different selections, the systematic uncertainty on total background rates appear to be very similar. In contrast the uncertainty of the signal rate seems more pronounced in the low mass selection, largely caused by jet energy scale uncertainty. Taking into account the limited statistics used in this study the search is found to be limited

by statistics and systematic uncertainties similarly, to the level of $\approx 20\%$ in each case. However, both types of uncertainty are expected to reduce with more data and when a deeper understanding of the detector performance is achieved. Final results were presented in the form of 95 % C.L. limits on the signal cross section relative to the SM prediction as a function of Higgs boson mass for an integrated luminosity of $33.4\,\mathrm{pb}^{-1}$. These results indicated that this analysis has the most sensitivity for Higgs masses in the 250 and $350\,\mathrm{GeV}/c^2$ range. Overall a similar level of performance is observed as in the GF analysis [41] when the order of magnitude lower cross section for the VBF signal is taken into account.

.

Chapter 8

Conclusions

The Higgs boson is the last remaining missing piece to the SM which has yet to be experimentally verified. The ATLAS experiment is one of two experiments at the LHC which is being used to try and find evidence of its existence. This year searches have intensified a great deal and have resulted in a large range of Higgs boson masses being excluded to 95 %C.L.

This study has made a first dedicated look into the sensitivity of the $H \rightarrow ZZ \rightarrow llll\nu\nu$ final state where the Higgs boson is produced by the VBF mechanism, for Higgs boson masses between m_H = 200-600 GeV/c^2. The analysis was performed using fully simulated Monte Carlo samples including trigger information. The performance of the reconstruction of electrons, muons and jets in the signal was shown to be robust against the main backgrounds. In order to take into account the changing signal properties with increasing Higgs mass, a low and high mass selection were developed, with the best performing analysis chosen as a function of Higgs mass. In the low mass selection, the lower E_T^{miss} levels in the signal mean that it relies more heavily on the unique properties of tag-jets associated with the VBF topology to suppress background. In contrast the high levels of E_T^{miss} in higher Higgs boson mass signals, was shown to give the most powerful discrimination over the background. An optimization procedure was performed on each of the selections and yielded of the order of a 10 % improvement in signal significance over the Higgs masses investigated relative to the baseline selections outlined.

The main backgrounds to both selections have been identified as $t\bar{t}$ and Z+jets, which together make up ≈80 % of the total background in both the low and high mass selections. Methods were explored to try and estimate their contribution using data driven techniques. The method used to estimate $t\bar{t}$ verified the uncertainty assumed on the $t\bar{t}$ cross section. The effect of various systematic uncertainties was taken into account and the combined systematic uncertainty predicted to be of the same order as statistical uncertainty, ≈20 %.

95 % C.L. limits on the signal cross section relative to the SM prediction as a function of Higgs boson mass showed that this analysis appears to be most sensitive to Higgs masses between 250 and 350 GeV/c^2. Taking into account its order of magnitude lower cross section, this analysis was shown to have a similar performance to the GF analysis [41].

Although large regions of the SM Higgs mass have been excluded to 95 %C.L. in recent times of LHC running, at the time of writing pockets of un-excluded Higgs mass remain in the mass region considered in this analysis. Although this would need to be re-evaluated using more data, the studies in this thesis have shown this channel could effectively contribute to this search.

References

[1] Particle Data Group Collaboration, K. Nakamura et al., *Review of particle physics*, J. Phys. **G37** (2010) 075021. 10, 54, 58, 101

[2] V. Barger and R. Phillips, *Collider Physics*. Updated Edition, Persus Publishing, Cambridge, Massachusetts, 1996. 11

[3] F. Halzen and A. D. Martin, *Quarks and Leptons: An Introductory Course in Modern Particle Physics*. John Wiley and Sons, 1984. 11

[4] P. Dirac, *The Quantum Theory of the Electron*, Proc. Royal. Soc.A117 **610624** (1928) . 12

[5] C.-N. Yang and R. L. Mills, *Conservation of isotopic spin and isotopic gauge invariance*, Phys. Rev. **96** (1954) 191–195. 12

[6] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. 10 **531533** (1963) . 13

[7] S. L. Glashow, *Partial symmetries of weak interactions*, Nucl.Phys.22 **579588** (1961) . 14

[8] A. Salam, *Weak and electromagnetic interactions*, Proc. of the 8th Nobel Symposium on Elementary particle theory, relativistic groups and analyticity **367377** (1968) . 14

[9] S. Weinberg, *A model of leptons*, Phys. Rev. Lett. 19 **12641266** (1967) . 14

[10] P. W. Higgs, *Broken Symmetries And The Masses Of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508–509. 14

[11] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321–322. 14

[12] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13** (1964) 585–587. 14

[13] A. Djouadi, *The Anatomy of electro-weak symmetry breaking. I: The Higgs boson in the standard model*, Phys. Rept. **457** (2008) 1–216, arXiv:hep-ph/0503172. 15, 25

[14] T. Hambye and K. Riesselmann, *Matching conditions and Higgs boson mass upper bounds reexamined*, Phys. Rev. D **55** (Jun, 1997) 7255–7262. 20

[15] The CDF and D0 Collaboration, *Combined CDF and D0 Upper Limits on Standard Model Higgs Boson Production with up to 8.6 fb-1 of Data*, arXiv:1107.5518 [hep-ex]. * Temporary entry *. 20, 21

[16] The ATLAS Collaboration, *Update of the Combination of Higgs Boson Searches in pp Collisions at $\sqrt{s}$=7 TeV with the ATLAS Experiment at the LHC*, ATLAS-CONF-2011-135 (2011) . 20, 21

[17] https://twiki.cern.ch/twiki/bin/view/CMSPublic/Hig11022TWiki. 20, 21

[18] ALEPH and CDF and D0 and DELPHI and L3 and OPAL and SLD and LEP Electroweak Working Group and Tevatron Electroweak Working Group and SLD Electroweak Working Group and Heavy Flavour Group Collaboration, *Precision Electroweak Measurements and Constraints on the Standard Model*, arXiv:0811.4682 [hep-ex]. 22

[19] LHC Higgs Cross Section Working Group, S. Dittmaier, et al., *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*, CERN-2011-002 (CERN, Geneva, 2011) , arXiv:1101.0593 [hep-ph]. 22, 54, 147, 148

[20] U. Aglietti et al., *Tevatron for LHC report: Higgs*, arXiv:hep-ph/0612172. 24

[21] O. S. Brüning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock, *LHC Design Report v.1 : The Main Ring, CERN-2004-003-V-1.* CERN, Geneva, 2004. 27

[22] O. S. Brüning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock, *LHC Design Report v.2 : The LHC Infrastructure and General Services, CERN-2004-003-V-2.* CERN, Geneva, 2004. 27

[23] M. Benedikt, P. Collier, V. Mertens, J. Poole, and K. Schindl, *LHC Design Report v.3 : The LHC Injector Chain, CERN-2004-003-V-3.* CERN, Geneva, 2004. 27

[24] L. Evans and P. Bryant, *LHC Machine*, Journal of Instrumentation **3** (2008) S08001. 27

[25] J. P. Blewett, *200 GeV intersecting storage accelerators*, Proceedings of the 8th International Conference on High-Energy Accelerators (1971) . 28

[26] The ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, Journal of Instrumentation 3, S08003 (2008) . 28, 32, 34, 35, 37, 38, 41

[27] S. Catani, *Aspects of QCD, from the Tevatron to the LHC*, arXiv:hep-ph/0005233. 30

[28] The ATLAS Collaboration, *ATLAS detector and physics performance: Technical Design Report*, ATLAS-TDR-014 (1999) . 31

[29] The ATLAS Collaboration, G. Aad et al., *Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics*, arXiv:0901.0512 [hep-ex]. 31, 61, 70

[30] The ATLAS Collaboration, *Magnet System Technical Design Report*, CERN/LHCC/97-18 (1997) . 33

[31] The ATLAS Collaboration, *Inner Detector Technical Design Report vol I*, CERN/LHCC/97-16 (1997) . 34

[32] The ATLAS Collaboration, *Inner Detector Technical Design Report vol II*, CERN/LHCC/97-17 (1997) . 34

[33] The ATLAS Collaboration, *Liquid Argon Technical Design Report*, CERN/LHCC/96-41 (1996) . 37

[34] K. Kleinknecht, *Particle Radiation*. Cambridge, 1998. 37

[35] The ATLAS Collaboration, *Tile Calorimeter Technical Design Report*, CERN-LHCC-96-12 (1996) . 39

[36] The ATLAS Collaboration, *Muon Spectrometer Technical Design Report*, CERN/LHCC/97-22 (1997) . 40

[37] P. Jenni, M. Nessi, M. Nordberg, and K. Smith, *ATLAS high-level trigger, data-acquisition and controls: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 2003. 42

[38] A. Gesualdi-Mello et al., *Overview of the High-Level Trigger Electron and Photon Selection for the ATLAS Experiment at the LHC*, IEEE Trans. Nucl. Sci. 5 28392843. (2002) . 43

[39] The ATLAS Collaboration, *ATLAS First-Level Trigger Technical Design Report*, CERN-LHCC-98-14 (1998) . 42

[40] S. Angelidakis et al., *Search for a Standard Model Higgs in the mass range 200-600 GeV in the channels $H \to ZZ \to llνν$ and $H \to ZZ \to llqq$*, Tech. Rep. ATL-COM-PHYS-2011-041, CERN, Geneva, Jan, 2011. 46, 58, 63, 66, 94, 95, 97, 101

[41] The ATLAS Collaboration, *Search for a Standard Model Higgs Boson in the Mass Range 200-600 GeV in the Channels $H \to ZZ \to llνν$ and $H \to ZZ \to llqq$ with the ATLAS Detector*, Tech. Rep. ATLAS-CONF-2011-026, CERN, Geneva, Mar, 2011. 46, 94, 97, 148, 152, 156, 162, 165, 167

[42] E. Gabrielli and B. Mele, *Testing Effective Yukawa Couplings in Higgs Searches at the Tevatron and LHC*, Phys.Rev. **D82** (2010) 113014, arXiv:1005.2498 [hep-ph]. 46

[43] M. A. Dobbs et al., *Les Houches guidebook to Monte Carlo generators for hadron collider physics*, arXiv:hep-ph/0403045. 53

[44] S. Frixione and B. R. Webber, *Matching NLO QCD computations and parton shower simulations*, JHEP **06** (2002) 029, arXiv:hep-ph/0204244. 53, 58

[45] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **05** (2006) 026, arXiv:hep-ph/0603175. 53, 54, 64

[46] P. Golonka and Z. Was, *PHOTOS Monte Carlo: A Precision tool for QED corrections in Z and W decays*, Eur. Phys. J. **C45** (2006) 97–107, arXiv:hep-ph/0506026. 54

[47] Z. Was, *TAUOLA the library for tau lepton decay, and KKMC/KORALB/KORALZ/... status report*, Nucl. Phys. Proc. Suppl. **98** (2001) 96–102, arXiv:hep-ph/0011305. 54

[48] J. M. Butterworth, J. R. Forshaw, and M. H. Seymour, *Multiparton interactions in photoproduction at HERA*, Z. Phys. **C72** (1996) 637–646, arXiv:hep-ph/9601371. 58

[49] The ATLAS Collaboration, *Measurement of the top quark-pair production cross section with ATLAS in pp collisions at $\sqrt{s}$=7 TeV*, Eur. Phys. J. **C71** (2011) 1577, arXiv:1012.1792 [hep-ex]. 58, 85

[50] S. Moch and P. Uwer, *Theoretical status and prospects for top-quark pair production at hadron colliders*, Phys. Rev. **D78** (2008) 034003, arXiv:0804.1476 [hep-ph]. 58

[51] U. Langenfeld, S. Moch, and P. Uwer, *New results for $t\bar{t}$ production at hadron colliders*, arXiv:0907.2527 [hep-ph]. 58

[52] M. Aliev et al., – *HATHOR – HAdronic Top and Heavy quarks crOss section calculatoR*, Comput. Phys. Commun. **182** (2011) 1034–1046, arXiv:1007.1327 [hep-ph]. 58

[53] J. Pumplin et al., *New generation of parton distributions with uncertainties from global QCD analysis*, JHEP **07** (2002) 012, arXiv:hep-ph/0201195. 58

[54] J. M. Campbell and R. K. Ellis, *An update on vector boson pair production at hadron colliders*, Phys. Rev. **D60** (1999) 113006, arXiv:hep-ph/9905386. 58, 59

[55] T. Binoth, N. Kauer, and P. Mertsch, *Gluon-induced QCD corrections to $pp \to ZZ \to \ell\bar{\ell}\ell'\bar{\ell}'$*, arXiv:0807.0024 [hep-ph]. 58

[56] The ATLAS Collaboration, *Measurement of the WW production cross section in proton-proton collisions at $s = 7$ TeV with the ATLAS detector*, Tech. Rep. ATLAS-CONF-2011-015, CERN, Geneva, Mar, 2011. 59

[57] The ATLAS Collaboration, *Measurement of the WW cross section in $\sqrt{s}$=7 TeV pp collisions with ATLAS*, arXiv:1104.5225 [hep-ex]. 59

[58] S. Allwood-Spires et al., *Monte Carlo samples used for top physics: Top Working Group Note IX*, Tech. Rep. ATL-COM-PHYS-2010-836, CERN, Geneva, Oct, 2010. 59

[59] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A. D. Polosa, *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, JHEP **07** (2003) 001, arXiv:hep-ph/0206293; hep-ph/0210213. 61, 63

[60] G. Corcella et al., *HERWIG 6.5: an event generator for Hadron Emission Reactions With Interfering Gluons (including supersymmetric processes)*, JHEP **01** (2001) 010, arXiv:hep-ph/0011363. 61, 63

[61] https://twiki.cern.ch/twiki/bin/view/AtlasProtected/TopMC2009QCD_Dijet_7_TeV. 64

[62] 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*, Tech. Rep. ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003, CERN, Geneva, Apr, 2008. 70

[63] M. Aharrouche et al., *Expected electron performance in the ATLAS experiment*, Tech. Rep. ATL-PHYS-INT-2010-126, CERN, Geneva, Nov, 2010. 72

[64] *ATLAS Egamma Combined Performance Group, Energy Scale and Resolution Recommendations*, https://twiki.cern.ch/twiki/bin/view/AtlasProtected/EnergyScaleResolutionRecommendations. 72, 152, 155

[65] *ATLAS Egamma Combined Performance Group, Efficiency Measurements*, https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/EfficiencyMeasurements. 73

[66] S. Hassani et al., *A muon identification and combined reconstruction procedure for the ATLAS detector at the LHC using the (MUONBOY, STACO, MuTag) reconstruction packages*, Nucl. Instrum. Meth. **A572** (2007) 77–79. 76, 77

[67] T. Lagouri et al., *A muon identification and combined reconstruction procedure for the ATLAS detector at the LHC at CERN*, IEEE Trans. Nucl. Sci. **51** (2004) 3030–3033. 76

[68] *Track Reconstruction in the ATLAS Muon Spectrometer with MOORE*, Tech. Rep. ATL-SOFT-2003-007, CERN, Geneva, Oct, 2003. 76

[69] R. Nicolaidou, L. Chevalier, S. Hassani, J. F. Laporte, E. L. Menedeu, and A. Ouraou, *Muon identification procedure for the ATLAS detector at the LHC using Muonboy reconstruction package and tests of its performance using cosmic rays and single beam data*, Journal of Physics: Conference Series **219** (2010) no. 3, 032052. http://stacks.iop.org/1742-6596/219/i=3/a=032052. 76

[70] K. Nikolopoulos, D. Fassouliotis, C. Kourkoumelis, and A. Poppleton, *Event-by-Event Estimate of Muon Energy Loss in ATLAS*, IEEE Transactions on Nuclear Science **54** (Oct., 2007) 1792–1796. 76

[71] T. Cornelissen, M. Elsing, S. Fleischmann, W. Liebig, E. Moyse, and A. Salzburger, *Concepts, Design and Implementation of the ATLAS New Tracking (NEWT)*, Tech. Rep. ATL-SOFT-PUB-2007-007. ATL-COM-SOFT-2007-002, CERN, Geneva, Mar, 2007. 76

[72] S. Tarem, Z. Tarem, N. Panikashvili, and O. Belkind, *MuGirl - Muon identification in ATLAS from the inside out*, Nuclear Science Symposium Conference Record, 2007 IEEE **1** (2007) 617–621. 77

[73] *ATLAS Muon Combined Performance Group, Guidelines for Analysis in Release 15*, https://twiki.cern.ch/twiki/bin/view/AtlasProtected/MCPAnalysisGuidelinesRel15. 78

[74] F. Cerutti, C. Gatti, P. Kluit, O. Kortner, W. Liebig, J. Liu, G. Salamanna, A. Salvucci, E. van der Poel, and J. Zhu, *Muon Momentum Resolution in First Pass Reconstruction of pp Collision Data Recorded by ATLAS in 2010*, Tech. Rep. ATLAS-COM-CONF-2011-003, CERN, Geneva, Jan, 2011. 79

[75] The ATLAS Collaboration, *Determination of the muon reconstruction efficiency in ATLAS at the Z resonance in proton-proton collisons at* $\sqrt{s}$*=7 TeV*, Tech. Rep. ATLAS-CONF-2011-008, CERN, Geneva, Feb, 2011. 79

[76] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-k_T jet clustering algorithm*, JHEP **04** (2008) 063, arXiv:0802.1189 [hep-ph]. 82

[77] The ATLAS Collaboration, *Measurement of inclusive jet and dijet cross sections in proton-proton collisions at 7 TeV centre-of-mass energy with the ATLAS detector*, arXiv:1009.5908 [hep-ex]. 82

[78] M.Cacciari, G.P.Salam, and G.Soyez. http://fastjet.fr/. 82

[79] The ATLAS Collaboration, B. Heinemann, F. Hirsch, and S. Strandberg, *Performance of the ATLAS Secondary Vertex b-tagging Algorithm in 7 TeV Collision Data*, Tech. Rep. ATLAS-COM-CONF-2010-042, CERN, Geneva, May, 2010. 85

[80] https://twiki.cern.ch/twiki/bin/view/AtlasProtected/HowToCleanJetsBad_jets_rel15_data. 85

[81] *Measurement of Missing Tranverse Energy*, Tech. Rep. ATL-PHYS-PUB-2009-016. ATL-COM-PHYS-2009-118, CERN, Geneva, Mar, 2009. Draft PUB Note for 'Measurement of Missing Transverse Energy in ATLAS' Chapter (J13). 86

[82] The ATLAS Collaboration, *Performance of the Missing Transverse Energy Reconstruction and Calibration in Proton-Proton Collisions at a Center-of-Mass Energy of 7 TeV with the ATLAS Detector*, Tech. Rep. ATLAS-CONF-2010-057, CERN, Geneva, Jul, 2010. 86

[83] K. Cranmer, Y. Fang, B. Mellado, S. Paganis, W. Quayle, et al., *Prospects for Higgs searches via VBF at the LHC with the ATLAS detector*, arXiv:hep-ph/0401148 [hep-ph]. 113

[84] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, The European Physical Journal C **71** (Feb, 2011) , arXiv:1007.1727v2. 131

[85] *Updated Luminosity Determination in pp Collisions at $\sqrt{s}$=7 TeV using the ATLAS Detector*, Tech. Rep. ATLAS-CONF-2011-011, CERN, Geneva, Mar, 2011. 152

[86] The ATLAS Collaboration, *Jet energy scale and its systematic uncertainty for jets produced in proton-proton collisions at $\sqrt{s}$=7 TeV and measured with the ATLAS detector*, Tech. Rep. ATLAS-CONF-2010-056, CERN, Geneva, Jul, 2010. 152

[87] The ATLAS Collaboration, *Update on the jet energy scale systematic uncertainty for jets produced in proton-proton collisions at $\sqrt{s}$=7 TeV measured with the ATLAS detector*, Tech. Rep. ATLAS-CONF-2011-007, CERN, Geneva, Feb, 2011. 152

[88] M. Bosman, L. Fiorini, C. Helsens, A. Juste, L. Mir, J. Nadal, and V. Vorwerk, *Weighting method to propagate heavy-flavor tagging calibrations and related uncertainties*, Tech. Rep. ATL-COM-PHYS-2010-331, CERN, Geneva, Jun, 2010. 157

[89] A. L. Read, *Presentation of search results: The CL(s) technique*, J. Phys. **G28** (2002) 2693–2704. 162