A Measurement of the Recently Discovered Higgs Boson in the Decay into Two Photons with Associated Jets, Using the ATLAS Detector at the LHC.

Robert James Bullimore Cantrill

Department of Physics

Royal Holloway, University of London

A thesis submitted to the University of London for the

Degree of Doctor of Philosophy

January 14, 2014

DECLARATION

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

RJCantrill 14th January 2014

Robert J B Cantrill

Abstract

A new boson has been discovered and measurements are under way using the 7 TeV and 8 TeV proton-proton collision data from the Large Hadron Collider to determine whether or not this is the Higgs boson as predicted by the Standard Model of Particle Physics (SM). Experimentally measuring the nature of this particle's couplings to other particles will help determine this. The Standard Model Higgs boson is expected to be produced by a variety of production mechanisms. The SM prediction is that the gluon-gluon fusion (ggF) and vector boson fusion (VBF) production mechanisms are the two production processes with the highest and second-highest, rates respectively.

This thesis concentrates on the study of the Higgs boson via its decay into two photons, which was one of the key discovery channels. Part of this analysis is to measure the ratio ($\mathfrak{R}$) of these rates using $13\,\text{fb}^{-1}$ of ATLAS $\sqrt{s} = 8\,\text{TeV}$ proton-proton collision data and determine if $\mathfrak{R}$ is consistent with the SM prediction.

Using the diphoton decay channel, events were selected to form a category of data events which is enriched in VBF events with little gluon-gluon fusion contamination. The selection procedure was optimised using a boosted decision tree (BDT) multivariate classifier. The distinguishing feature of this analysis was that the BDT was trained using background events from the data sample, so as to reduce the dependency on the modelling of the background processes. It was shown that using a BDT classifier, the VBF signal significance improves by 24.0% relative to the standard cut-based analysis and suffers from 12.0% less ggF signal contamination. Using this event classification $\mathfrak{R}$ was measured as

$$\mathfrak{R} = \sigma_{VBF}/(\sigma_{ggF} + \sigma_{VBF}) = 0.037 \pm 0.067(\text{stat}) \pm 0.035(\text{syst})$$

where σ_{VBF} and σ_{ggF} are the respective cross sections of the vector boson fusion process and the gluon-gluon fusion process. The SM prediction is $\mathfrak{R} = 0.075$ Although the uncertainty on the current measurement is large, it is shown using pseudodata, that this choice of categorisation will help reduce the uncertainty on $\mathfrak{R}$ when more data are available.

Acknowledgements

I'd like to take the opportunity to thank the various people who have assisted me throughout the duration of my PhD. Without my supervisor Dr Pedro Teixeira-Dias, I could not have completed this project. He has provided me with quality advice and much encouragement over the past four years. I would also like to thank all the other members of the Royal Holloway particle physics group for their support and collaboration, especially to Dr Ricardo Gonçalo, for his supervision whilst I was based at CERN.

Thanks to all my friends at Royal Holloway and CERN for keeping me sane over the past four years. One particular friend I'd like to thank is Tim Brooks. Before I began my PhD, my computing knowledge was extremely limited and Tim gave up a lot of his time to help me get started.

The past four years have been an amazing experience and it couldn't have been timed more perfectly. I've witnessed many important milestones in the LHC project, including the discovery of the new boson believed to be the elusive Higgs. I feel extremely privileged to be involved and to work with such amazing people on this very historical analysis.

This PhD was funded by the Science and Technology Facilities Council and relies on analysis of data provided by the ATLAS collaboration.

To my family:

Tim and Angela Cantrill,

Sarah, Stuart, Eliza and George Milnes.

Thank you for all your love and support you have shown me over the years.

I love you all very much.

Contents

List of Figures

List of Tables

Preface

The LHC has achieved many milestones since it started running in 2009. In July 2012, a new scalar boson was discovered by the ATLAS and CMS collaborations. Since then more data has been acquired and enough measurements have been made to suggest that the new particle is a Higgs boson. This new particle will be referred to as a Higgs boson throughout this thesis.

My work on the ATLAS experiment contributed to the search for the Higgs boson, including validating the $H \to \gamma\gamma$ analysis and to determine the systematic uncertainty due to the jet energy scale for the discovery analysis.

In the search for a Higgs boson signal in the diphoton final state, candidate events are grouped together in categories for the purpose of improving the signal sensitivity. One of these categories was developed to be enriched in signal events where the Higgs boson is produced via the vector boson fusion (VBF) process. A key feature of the VBF signal events is the production of forward jets, therefore the jet energy scale is an important systematic uncertainty; fluctuations in the jet energy measurements can potentially lead to events migrating between different event categories. I have contributed a description of the methodology for determining this systematic uncertainty in internal ATLAS communications [1, 2, 3, 4, 5, 6, 7], and the uncertainties I have calculated were used in public analyses [8, 9].

After the discovery, my work focused on defining, and then optimising, a new category enriched in VBF signal events and developing an associated stand-alone diphoton channel method to measure the ratio of the Higgs boson event production in the VBF and gluon-gluon fusion modes, using $13\,\text{fb}^{-1}$ of $8\,\text{TeV}$ ATLAS data collected in 2012. A measurement of this type could provide useful information to check compatibility with the Standard Model (SM) hypothesis, and in comparison with fits to the Standard Model hypothesis, could have increased the sensitivety to non-standard physics that would manifest itself only via $H \to \gamma\gamma$ decay loops. This unique work is

the main content of this thesis as described in Chapters 5-8.

In addition to my Higgs analysis work, I also undertook work in the trigger and the data acquisition systems. This included ways to improve the identification of electron candidates in the level 2 trigger, to improve usage of CPU time, memory allocation and disk space; and assisting in the programme of the rolling replacement of the Readout System (ROS) PCs in the USA15 cavarn, that was carried out in the latter half of 2011 to upgrade the ROS performance.

The thesis is structured as follows:

- In Chapter 1, the theory is outlined and the Mass Mechanism is explained. The Standard Model prediction of how the Higgs boson should behave is given, and recent measurements testing consistency with the prediction are highlighted.

- In Chapter 2, a brief description of the LHC and ATLAS detector is given. Design features relevant to this analysis are discussed.

- In Chapter 3, the signal and background processes that are modelled and studied in this analysis are discussed. The signal Monte Carlo samples are identified and the motivation to use data-driven background modelling is justified in this chapter.

- In Chapter 4, the procedure to reconstruct the photons, jets, electrons and muons using the measurements from the ATLAS detector is given.

- In Chapter 5, the event selection for the $H \rightarrow \gamma\gamma$ analysis is given. A categorisation procedure to separate $H \rightarrow \gamma\gamma$ events that are produced by different mechanisms is also shown in this chapter and the limitations with the cut-based categorisation procedure are identified. A classification using a boosted decision tree is investigated as an alternative.

- In Chapter 6, shows how the categorisation can be used to calculate the ratio between the vector boson fusion and the gluon-gluon fusion cross sections by utilising background fits and pseudodata.

- In Chapter 7, the measured result on the ratio of the vector boson fusion cross sections and the gluon-gluon fusion cross sections is presented using a new choice of event categorisation.

- In Chapter 8, the main systematic effects are explored.

- In Chapter 9, the conclusions are presented.

The material in Chapters 5-9 is my own work and Chapters 1-4 is a review of the literature to provide the background information relevant for this analysis, where information has been derived from other sources, it is cited this in the thesis.

Throughout this thesis natural units are used

$$c = \hbar = 1.$$

Energy, momentum and mass are given in electron-volts (eV).

Chapter 1

Theory and Motivation

1.1 Building the Standard Model of Particle Physics

The Standard Model of Particle Physics (SM) is currently the best description of the nature of fundamental particles and how these particles interact with one another. With the discovery of a new scalar boson and recent measurements suggesting that this new particle has the properties of a SM Higgs boson [10] the theory is practically complete.

The theory describes two basic types of particle: fermions and bosons. The fundamental fermions are quarks and leptons, often associated with matter. The bosons are the force-carrying particles which are exchanged by fermions interacting with one another. These particles are shown in Table 1.1. All processes are described by a renormalisable quantum field theory, which is invariant under gauge transformations [11]. It is convenient to use a Lagrangian formalism in this theory.

The importance of local gauge invariance is that, fields describing spin-1 vector particles can be introduced into the theory that leaves it renormalisable. In the SM the photon, W^+ boson, W^- boson, and Z boson arise from enforcing $SU(2)_L \times U(1)_Y$ local gauge invariance. There are three fields associated with the left-hand $SU(2)_L$ group:

$$\{W_\mu^1, W_\mu^2, W_\mu^3\} \in SU(2)_L$$

and a so-called hypercharge (Y) field, B_μ associated with the $U(1)_Y$ group. The W_μ^1 and the W_μ^2 combine to form the W^+ and the W^- boson associated with the nuclear weak interactions, as

illustrated in Figure 1.1(a). These interactions are parity violating, so for this reason the fermions are arranged as left-handed chiral doublets and right-handed singlets for each generation of quark and lepton[1].

$$X_{L,quarks} = \begin{pmatrix} u \\ d \end{pmatrix}_L, \quad \begin{pmatrix} c \\ s \end{pmatrix}_L, \quad \begin{pmatrix} t \\ b \end{pmatrix}_L$$
$$X_{L,leptons} = \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \quad \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L, \quad \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L$$

$$X_{R,quarks} = u_R, d_R, c_R, s_R, t_R, b_R$$
$$X_{R,leptons} = e_R, \mu_R, \tau_R$$

The W_μ^3 field, orthogonally mixes with the B_μ field as follows:

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B_\mu^Y \\ W_\mu^3 \end{pmatrix} \tag{1.1}$$

where θ_W is the mixing angle. The combination of these two fields form the electromagnetic photon field (A_μ) and the weak neutral Z field. Illustrated examples in the form of Feynman diagrams are shown in Figure 1.1(b) and Figure 1.1(c) respectively.

In summary, the electromagnetic and weak interactions are unified in a $SU(2)_L \times U(1)_Y$ group and the associated bosons can be described without destroying the renormalisabilty of the theory [12]. Unfortunately, the bosons that are described by local gauge invariance are massless and it is known from experiment that the W and Z bosons have mass $O(100)$ GeV. Adding the mass terms explicitly cannot be done as this will destroy the symmetries. A mass mechanism was therefore devised by Higgs, Kibble, Englert, Brout, Hagen and Guralnik [13, 14, 15, 16, 17, 18] to include these mass terms.

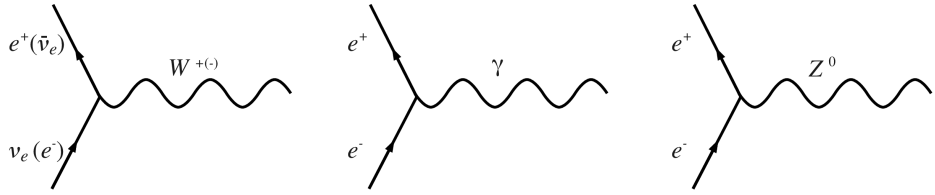
1.2 The Mass Mechanism

The intention is to generate mass terms for the W^+, W^- and Z^0 bosons in an $SU(2)_L \times U(1)_Y$ invariant Lagrangian that also incorporates a massless photon. This requires at least three additional

[1] In the SM the neutrino is assumed to be massless so there is no right-handed neutrino but a right handed anti-neutrino.

	Fermions			Bosons
Quarks	**Up** (u) Mass = 2.3 MeV Charge = $\frac{2}{3}e$ Spin = $\frac{1}{2}$	**Charm** (c) Mass = 1.28 GeV Charge = $\frac{2}{3}e$ Spin = $\frac{1}{2}$	**Top** (t) Mass = 173.5 GeV Charge = $\frac{2}{3}e$ Spin = $\frac{1}{2}$	**Gluon** (g) Mass = 0 eV Charge = $0e$ Spin = 1
	Down (d) Mass = 4.8 MeV Charge = $-\frac{1}{3}e$ Spin = $\frac{1}{2}$	**Strange** (s) Mass = 95 MeV Charge = $-\frac{1}{3}e$ Spin = $\frac{1}{2}$	**Bottom** (b) Mass = 4.18 GeV Charge = $-\frac{1}{3}e$ Spin = $\frac{1}{2}$	**W boson** (W) Mass = 80.4 GeV Charge = $\pm 1e$ Spin = 1
Leptons	**Electron** (ν_e) **Neutrino** Mass < 3 eV Charge = $0e$ Spin = $\frac{1}{2}$	**Muon** (ν_μ) **Neutrino** Mass < 0.19 MeV Charge = $0e$ Spin = $\frac{1}{2}$	**Tauon** (ν_τ) **Neutrino** Mass < 18.2 MeV Charge = $0e$ Spin = $\frac{1}{2}$	**Z Boson** (Z) Mass = 91.2 GeV Charge = $0e$ Spin = 1
	Electron (e) Mass = 0.511 MeV Charge = $-1e$ Spin = $\frac{1}{2}$	**Muon** (μ) Mass = 105.7 MeV Charge = $-1e$ Spin = $\frac{1}{2}$	**Tauon** (τ) Mass = 1.777 GeV Charge = $-1e$ Spin = $\frac{1}{2}$	**Photon** (γ) Mass = 0 eV Charge = $0e$ Spin = 1
				Higgs Boson (H) Mass 125.5 GeV Charge = $0e$ Spin = 0

Table 1.1: All particles in standard model. Fermions have anti-matter counter parts, which have opposite charge but the exact same mass. The values of the masses quoted are obtained from the particle data group [19]. The mass quoted for the Higgs boson mass is the mass measured by ATLAS (see Section 1.5.1).



(a) Charged Weak current (b) EM current (c) Neutral Weak current

Figure 1.1: Feynman diagrams showing examples of gauge interactions: (a) electromagnetic interaction: an electron emitting a photon. (b) charged weak interaction of an electron and a neutrino with a W boson. (c) weak neutral current of an electron emitting a Z boson.

degrees of freedom to be introduced into the model to provide the longitudinal polarisation modes for the massive gauge bosons. The simplest way in which to do this, is to introduce a complex doublet of fields, Φ, into the Lagrangian which is made up of 4 scalar electrically neutral and charged fields:

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{1.2}$$

with an associated potential of the form

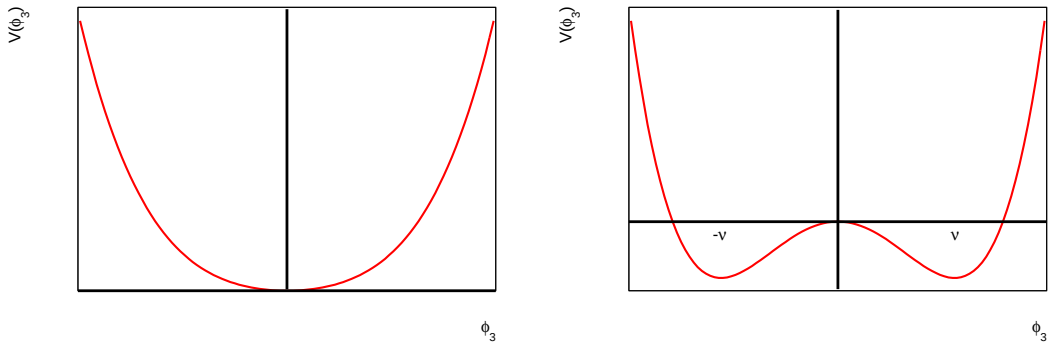
$$V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2 \tag{1.3}$$

For the case where $\mu^2 < 0$ and $\lambda > 0$ the shape of the potential is as shown in Figure 1.2. The shape of this potential has rotational symmetry but the minimum energy state is not zero at $\Phi^\dagger \Phi = 0$ as this would be unstable. The minimum energy state has a nonzero expectation value, v.

$$\Phi^\dagger \Phi = -\frac{\mu^2}{2\lambda} \equiv \frac{v^2}{2} \tag{1.4}$$

To generate the masses of the gauge bosons, a minimum is chosen such that $\phi_1 = \phi_2 = \phi_4 = 0$ and $\phi_3 = v$. Since the potential is invariant under rotational symmetry no generality is lost in the choice of the minimum, but in doing so the symmetry is spontaneously broken. The consequence of this choice gives rise to mass terms for the Z and W bosons. The mass of the W boson is

$$m_W = \frac{1}{2} v g \tag{1.5}$$



(a) $\mu^2 > 0, \lambda > 0$ (b) $\mu^2 < 0, \lambda > 0$

Figure 1.2: Schematic of the energy potential associated with the mass mechanism. (a) shows the shape where λ is positive and (b) shows the shape where λ is negative, in this case ϕ is non zero at the potential minima.

and the mass of the Z boson is

$$m_Z = \frac{1}{2} v \sqrt{g^2 + g'^2} \tag{1.6}$$

where g and g' are the $SU(2)_L$ and hypercharge coupling strengths respectively. It can be shown the two masses are related by the mixing angle

$$\frac{m_W}{m_Z} = \cos\theta_W . \tag{1.7}$$

In addition, the choice of the ϕ_1 and ϕ_2 having no vacuum expectation value allows the photon to remain massless. Higgs showed that an additional scalar boson is also predicted from this mechanism [14]. This became known as the Higgs boson, which has a mass m_H of

$$m_H^2 = 2v^2\lambda \tag{1.8}$$

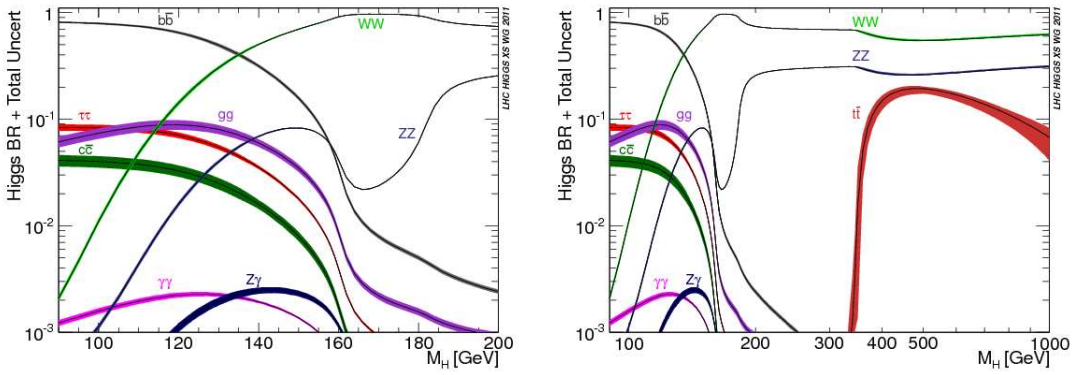
The choice of the vacuum expectation value also quantifies the strength of the Higgs boson couplings to the heavy gauge bosons and self interactions. Trilinear couplings of the Higgs boson with the other gauge bosons, show that the strength of the coupling is proportional to the masses of the gauge bosons [20].

In addition, the Higgs mechanism also provides an explanation for the fermion masses and associated coupling strengths to the Higgs boson. These appear in the "Yukawa terms" of the Lagrangian after the symmetry is broken. Although this term does not predict the masses of each

fermion, it predicts that the strength of each fermion coupling to the Higgs boson is proportional to the fermion mass. As the top quark is by far the heaviest SM fermion, this plays a huge role in Higgs interactions, as will be discussed later.

1.3 Higgs Boson Branching Ratios and Production Cross Sections

The Higgs boson is detected through its decay into other particles. Using the coupling information, it is known that the direct decays are $H \to WW$, $H \to ZZ$ and $H \to ff$. Although the mass of the Higgs boson is not predicted by the SM, the rate of each decay can be predicted for a given Higgs boson mass from knowing the coupling strengths and kinematic states. These are referred to as branching ratios and are shown in Figure 1.3. Although the top quark, the Z boson and the W boson are the heaviest known particles, for a light Higgs boson, $O(100)$ GeV, the dominant decay is $H \to b\bar{b}$. The top quark, the Z boson and the W boson are too heavy for a light Higgs boson to decay to and will only decay into these particles if they are off-shell.



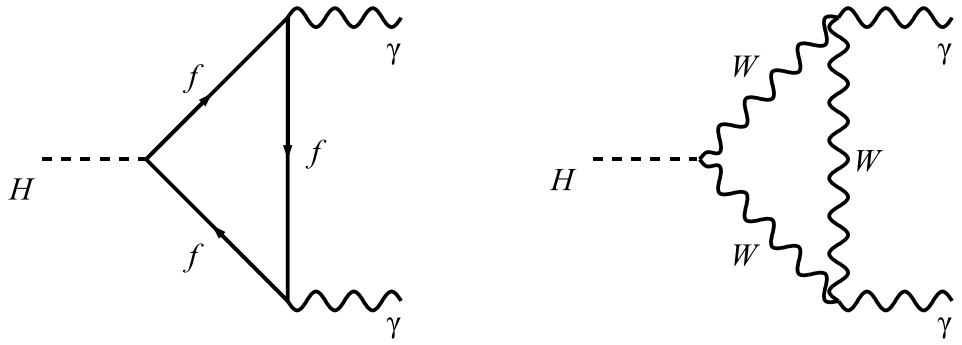
(a) Low mass (b) High mass

Figure 1.3: Branching ratios of the SM Higgs Boson. The branching ratios are shown for a range of possible SM Higgs boson masses. (a) The branching ratios for a light Higgs mass and (b) are the branching ratios for a Higgs mass up to 1 TeV [21].

It is also possible for a light mass Higgs boson to rarely decay into gluons and photons. Although these particles are massless, the decay process is allowed through loops of heavy particles. An example is shown in Figure 1.4 for the diphoton decay.

At the LHC there are five Higgs boson production mechanisms:

- Gluon-gluon fusion;



(a) $H \to \gamma\gamma$ mediated by a heavy fermion loop. (b) $H \to \gamma\gamma$ mediated by a loop of heavy gauge bosons.

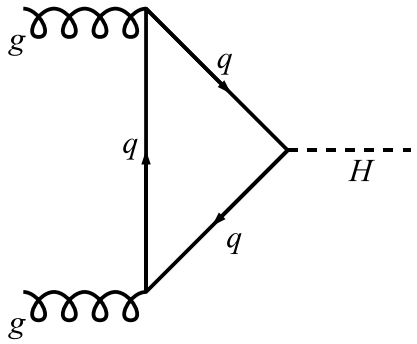
Figure 1.4: Feynman diagrams of the $H \to \gamma\gamma$ decay channel in the SM. The photons are massless, so the decay is mediated by heavy particle loops, which can either be heavy fermions, shown in (a) or massive gauge bosons, shown in (b).

- Vector boson fusion (VBF);
- Associated Higgs boson production with a W boson;
- Associated Higgs boson production with a Z boson;
- Associated Higgs boson production with a pair of top quarks.

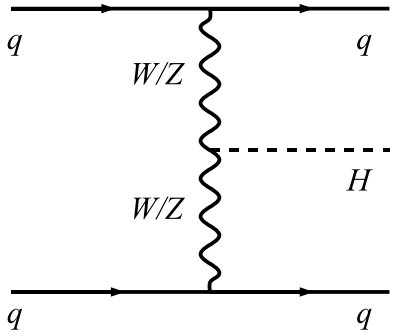
The leading order Feynman diagrams for these processes are shown in Figure 1.5. The cross section of each process is dependent on the mass of the Higgs boson, and has a trend to generally decrease with mass (see Figure 1.6).

The LHC is a hadron collider with a high centre of mass energy, therefore Higgs boson production via gluon-gluon fusion is the highest rate production process for a light Higgs boson as seen in Figure 1.6. Since gluons are massless, the gluon-gluon fusion process is mediated by a heavy quark loop. This is usually the top quark as it is by far the heaviest, and therefore has the strongest coupling to the Higgs boson. As this is a strongly interacting process, the cross section is modified substantially by radiative corrections. Higher order diagrams than the ones shown in Figure 1.5(a) have to be taken into consideration when calculating the cross section.

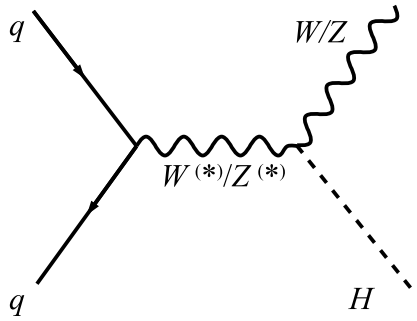
The VBF process is another common production process at the LHC. As seen in Figure 1.5(b), the Higgs boson is produced from weak bosons that are radiated out from a quark in each proton. To produce a Higgs boson, the energies required of the weak bosons have to be of the order of a Higgs boson mass, therefore the quarks carry away the majority of the energy. The transverse



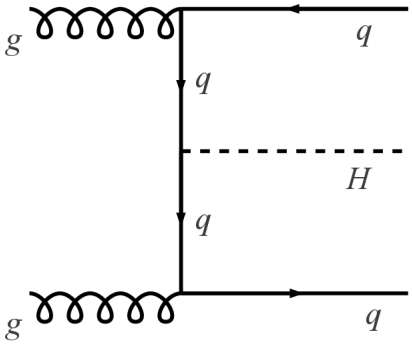
(a) gluon-gluon fusion



(b) vector boson fusion



(c) Associated production with a W or Z boson



(d) Associated production with two top quarks

Figure 1.5: Feynman diagrams of the five main production mechanisms. (a) Gluon-gluon fusion, (b) vector boson fusion, (c) associated production with a W or a Z boson and (d) associated production with a top and an anti-top quark.

momenta of the quarks is also large but much less than the total energy carried away. This means the recoiling quarks have a small scattering angle. In addition, decay products of the Higgs boson will be fairly centralised so the quarks and decay products will be largely separated.

The associated production mechanisms are an order of magnitude less than the VBF process. Although the cross section is very small, searching for the Higgs boson in association with an additional final state particle will enhance the signal sensitivity because there are fewer background processes with the same final state products.

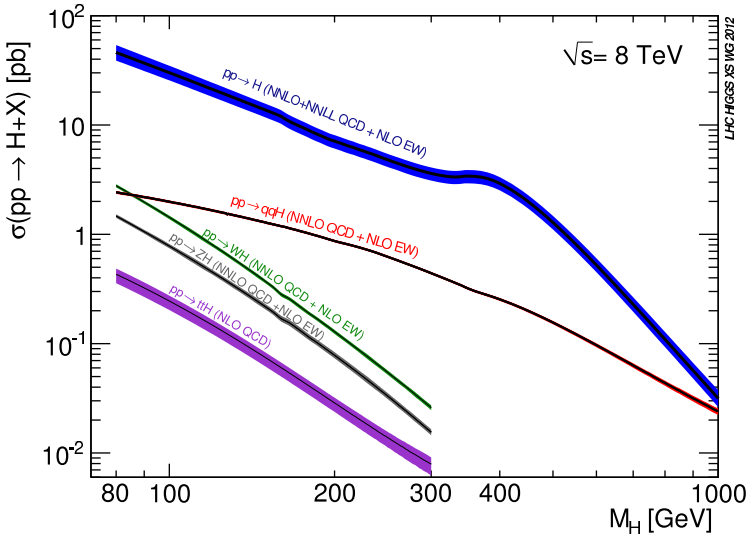


Figure 1.6: Cross sections of the SM Higgs Boson production mechanisms with feasible detection is shown for all possible SM Higgs boson masses [21].

1.4 Theoretical and Experiment Constraints on the Higgs Mass

In the SM the mass of the Higgs boson is not predicted. However in order for the Higgs mechanism to be valid in the SM, the mass of the Higgs boson has to be constrained to a particular mass range. The absolute upper mass limit allowed by the SM can be inferred from WW scattering. When all diagrams are taken into account for this process (two of which are mediated by the Higgs boson), the mass of the Higgs boson is restricted to $< O(1)$ TeV or else unitarity of the quantum scattering amplitude is violated.

Further constraints are also provided by electroweak corrections. As shown earlier the masses of the W and Z bosons are related by the mixing angle

$$\frac{M_W}{M_Z \cos\theta_W} = 1 \tag{1.9}$$

However, this is only at leading order, when radiative corrections (such as the ones shown in Figure 1.7) are taken into account, there are deviations from unity, which are dependent on the mass of the Higgs boson and the masses of the other particles involved in Figure 1.7. Experimental measurements of the other parameters, therefore help constrain the mass of the SM Higgs boson. The overall fit is shown in Figure 1.8, which predicts the Higgs mass to be no greater than 260 GeV with 95% confidence [22, 20].

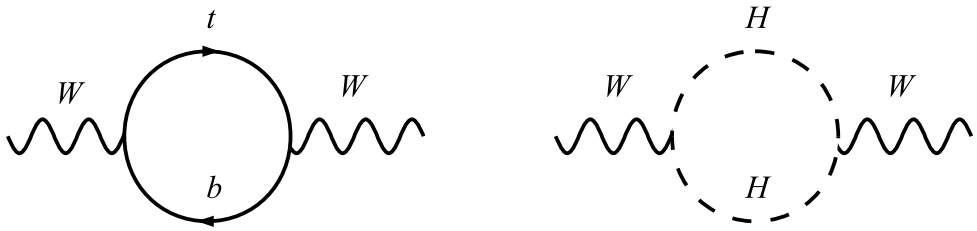


Figure 1.7: Radiative corrections to the W boson mass at the electroweak scale.

By the end of summer 2012, the direct searches from ATLAS had excluded the existence of a SM Higgs boson with a mass between 111 and 122 GeV and 131 and 559 GeV at over 95% confidence (see Figure 1.9(a)). An excess consistent with the SM hypothesis was observed at ≈ 125 GeV for both ATLAS and CMS with a significance of $6.0\,\sigma$ for ATLAS (see Figure 1.9(b)) and $5.0\,\sigma$ for CMS.

1.5 Measurements of the Higgs Boson

At present the ATLAS and CMS LHC experiments have observed a new particle, which has been observed in five decay channels $H \to \gamma\gamma$, $H \to ZZ$, $H \to WW$, $H \to \tau\tau$ and $H \to b\bar{b}$. The properties of the new particle have been measured to check for consistency with the SM Higgs boson.

1.5.1 Mass Measurement

As the $H \to \gamma\gamma$ and $H \to ZZ \to 4$ leptons ($4e$, 4μ or $2e\,2\mu$) decay channels produce a reconstructable mass peak (see Figure 1.10) these channels are used to obtain a mass measurement. With the available data the combined Higgs mass measurement is

$$m_H = 125.5 \pm 0.2(\text{stat})^{+0.5}_{-0.6}(\text{syst})\,\text{GeV}$$

from ATLAS [10] and

$$m_H = 125.7 \pm 0.4(\text{stat}) \pm 0.3(\text{syst})\,\text{GeV}$$

from CMS [23].

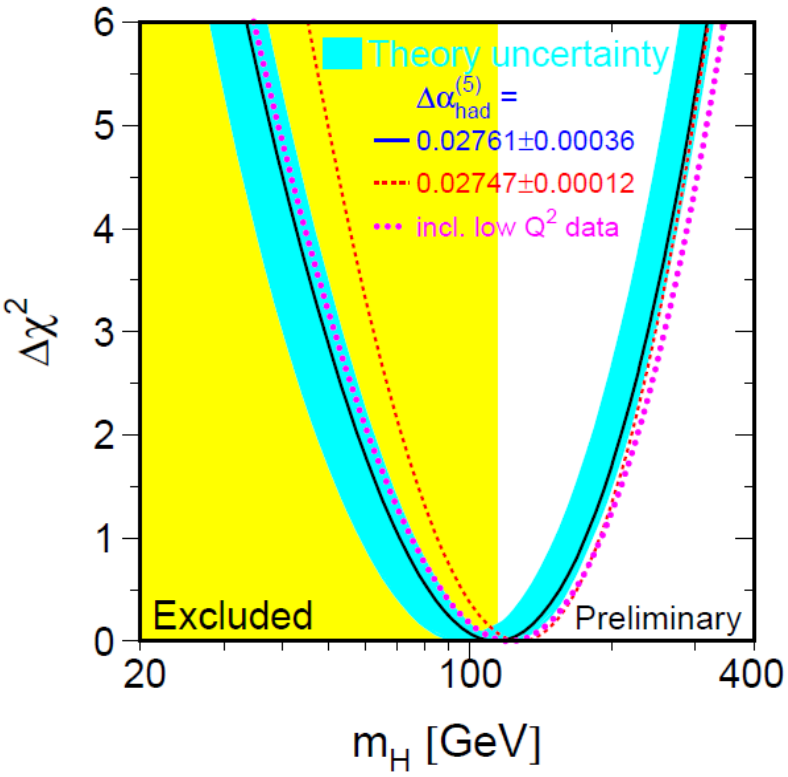
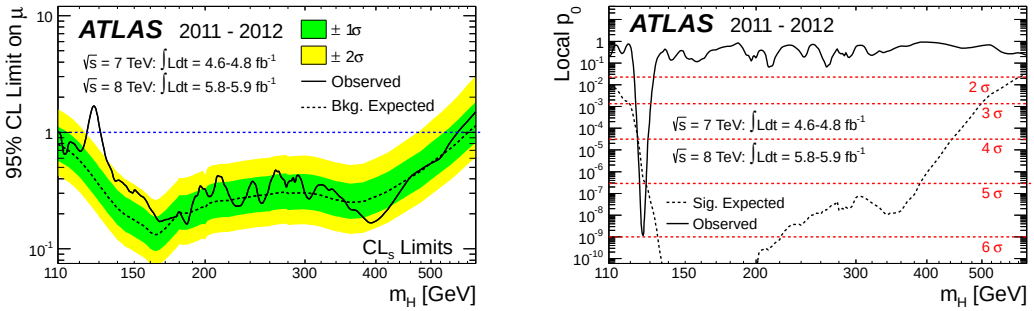


Figure 1.8: Global fit of electroweak parameters which predict a likely value of the SM Higgs mass, assuming validity of the SM. The Large Electron Positron (LEP) collider, excluded a SM Higgs boson with a mass up to 114 GeV with 95% confidence. The SM fit constrains the SM Higgs mass to be lower than 260 GeV with 95% confidence level [22].

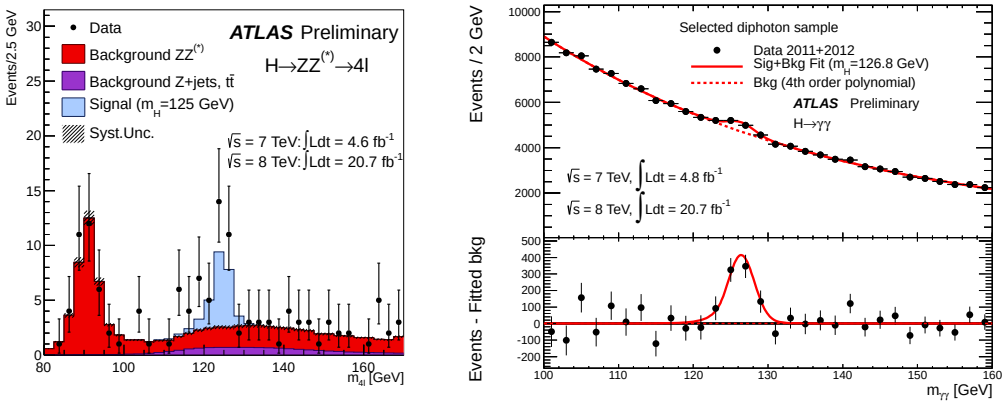
1.5.2 Couplings to the Decay Particles

If the observed particle is the SM Higgs boson the various decay channels are predicted to occur at the rates shown previously in Figure 1.3. A signal strength parameter, μ, is defined which measures the rate of decay for a given decay process relative to the SM prediction. An observation compatible with the background-only hypothesis corresponds to $\mu = 0$. An observation consistent with the SM signal hypothesis corresponds to $\mu = 1$. If $\mu > 1$ the decay occurs more often than the SM prediction. The recent measurements of μ from ATLAS and CMS are shown in Figure 1.11. The overall measurement of μ from both experiments is comparable with the SM prediction within the current experimental uncertainties.



(a) 95% Confidence exclusion limits (b) Signal Significance

Figure 1.9: Results from 4.6-4.8 fb^{-1} of 7 TeV data and 5.8-5.9 fb^{-1} of 8 TeV ATLAS data[8]. (a) The SM Higgs boson with a mass between 111 and 122 GeV and 131 and 559 GeV has been excluded with at least 95% confidence. (b) Signal significance and the probability of obtaining an excess for a given mass, assuming the background only hypothesis.

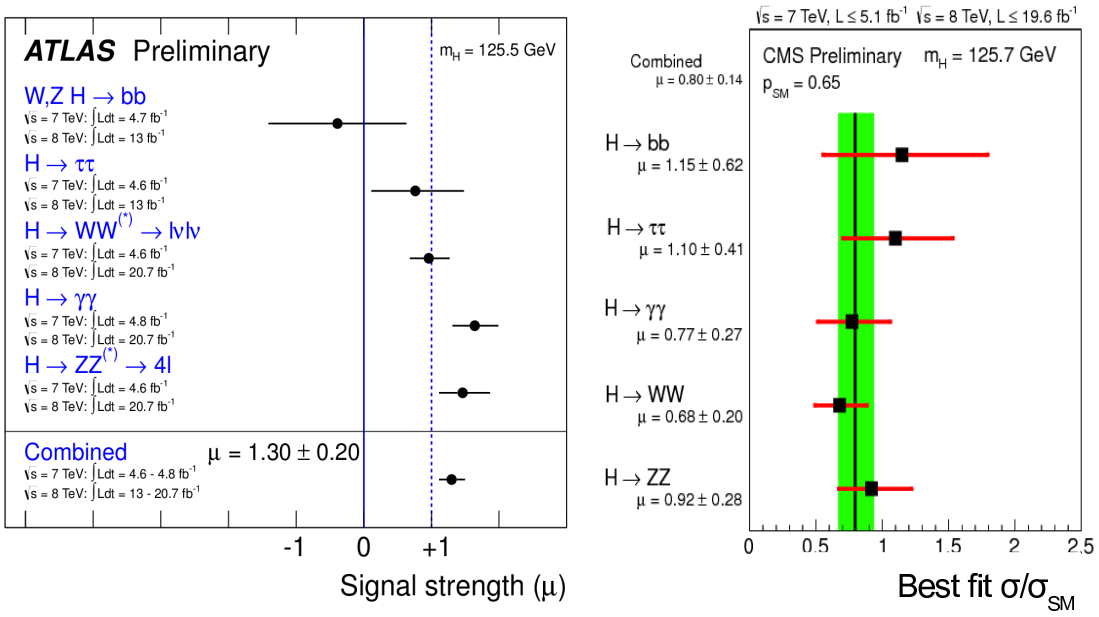


(a) H→ ZZ channel (b) H→ $\gamma\gamma$ channel

Figure 1.10: Invariant mass distributions using 25 fb^{-1} of ATLAS data shown for two of the Higgs boson search channels. (a) $H \to \gamma\gamma$ [24] and (b) $H \to ZZ \to 4$ leptons [25].

1.5.3 Spin and Parity

The spin and parity of the new boson have also been measured with the current data. If the particle is the SM Higgs boson, its spin should be zero and the parity should be even. The spin and parity of the particle have been measured using various decay angular distributions of the final state particles in the selected events. The observation of the decay into two photons automatically implies that this particle is not a vector boson of spin 1. Using the $H \to \gamma\gamma$, $H \to WW \to l\nu l\nu$ and H→ $ZZ \to ll$ channels the ATLAS and CMS data exclude a spin 2 CP odd particle with over 99% confidence and favours spin 0 CP even [28].



(a) ATLAS signal strength

(b) CMS signal strength

Figure 1.11: Signal strength determined for five decay processes, signal hypothesis shown by the dashed lines. $\mu = 1.30$ calculated by ATLAS at $m_H = 125.5\,\text{GeV}$ in (a) [26] and $\mu = 0.80$ calculated by CMS at $m_H = 125.7\,\text{GeV}$ in (b) [27].

Measurements of all these properties give strong evidence that the newly observed particle is consistent with the SM Higgs boson.

1.5.4 Production Cross Sections

The measurements of the cross section for the various production processes of this new particle are also important to further test consistency with the hypothesis that it is a Higgs boson, or even the SM Higgs boson. Two particular processes of interest are the gluon-gluon fusion and VBF as these cross sections are high enough for feasible detection. The gluon-gluon fusion process provides a measurement of the Higgs boson coupling to fermions due to the quark loop, and the VBF process provides a measurement of the Higgs boson couplings to the W and Z bosons. The SM predicts the gluon-gluon fusion process to occur 12.3 times more often than VBF assuming the Higgs mass $m_H = 125.5\,\text{GeV}$ [21].

The gluon-gluon fusion process is mostly mediated by a the top quark loop. This is because the strength of the Higgs boson coupling to fermions is directly proportional to their mass. There are however models that go beyond the SM. For example supersymmetric extensions of the SM

(SUSY), predict the existence of fermionic particles for every bosonic particle and bosonic particles for every fermionic particle. Assuming there are additional heavy SUSY fermions in nature, in addition to the top quark, there could be additional particles in the gluon-gluon fusion loop which could enhance the gluon-gluon fusion cross section.

Other models, with reduced or suppressed Higgs boson couplings to fermions or bosons, could result in a significant reduction of either the gluon-gluon fusion process or the VBF process. A direct measurement of the ratio of the gluon-gluon fusion and the VBF productions could provide useful information to check for compatibility with the SM Higgs hypothesis or otherwise.

Chapter 2

The LHC and ATLAS Detector

In this chapter an overview of the Large Hadron Collider (LHC) and the ATLAS experiment is given. The performance of the ATLAS detector in relation to the $H \rightarrow \gamma\gamma$ signal and the pp collision data will be the main focus.

2.1 The LHC

The LHC is a high energy particle collider at the CERN laboratory in Geneva. The aim of the LHC is to uncover new physics at high energies by accelerating two beams of particles in opposite directions, in a ring and colliding them together at fixed points. Since the start of the LHC operation there have been several physics programmes to study two types of collision: proton on proton (pp) and heavy ion collisions.

In 2012, proton beams were accelerated to energies of 4 TeV each, creating a centre of mass energy of 8 TeV. Proton beams are accelerated in bunches of $\approx 10^{11}$ protons by radio frequency (RF) acceleration cavities and steered round the ring by powerful superconducting magnets. As many as 1400 bunches are present in one beam making the bunch-crossing rate extremely high. Each beam travels in a high vacuum beampipe, to reduce collisions with molecules, maintaining the beam lifetime. At each collision point the bunches are focused and squeezed by powerful quadrupole magnets. During a bunch crossing there is likely to be a pp interactions or collisions, which are measured by sophisticated particle detectors. Each detector is designed for specific types of analysis in the hope of new physics discoveries. There are two general-purpose experiments: ATLAS and CMS, both designed for a multitude of physics searches and studies. ALICE

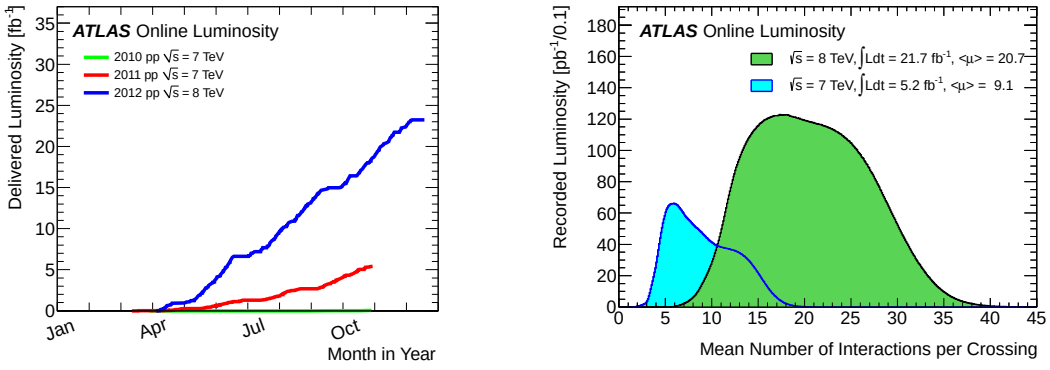
is designed to study heavy ion collisions during the heavy ion runs and LHCb is designed to study collision events containing B mesons [29].

2.1.1 LHC Performance

The number of data events of a particular process (N_p) collected is measured by the integrated luminosity ($\mathcal{L}$)

$$N_p = \sigma_p \mathcal{L} = \sigma_p \int L dt \tag{2.1}$$

where σ_p is the cross section of a particular process and L is the instantaneous luminosity. As the beam is not continuously replenished, during a run the instantaneous luminosity will decay. When the instantaneous luminosity becomes too low the beams are dumped and the LHC is refilled to start a new run. Throughout 2012, the LHC was operating at a very high instantaneous luminosity, making the 2012 dataset the largest of the total LHC operation so far, as shown in Figure 2.1(a). The data taking rate in 2012 was better than previous years due to larger number of bunches per beam, better control of the beam and quicker turn around periods between each run. However this also meant that pp interactions were occurring at greater rates. During 2012 there were more interactions per bunch crossing on average compared with 2011 data as shown in Figure 2.1(b). When there is more than one interaction per bunch crossing, this is referred to as pileup, which can affect event reconstruction and analysis procedures are in place to get around this.



(a) (b)

Figure 2.1: (a) Amount of data recorded in fb $^{-1}$ during 2010, 2011 and 2012 shown separately. (b) Differential integrated luminosity with respect to the average number of interactions per bunch crossing [30].

2.2 ATLAS

The ATLAS detector [29] is a general purpose detector composed of several layers of subdetectors. The subdetectors are arranged in concentric layers around the beam axis (referred to as the barrel) and in flat layers either side of the barrel (referred to as the end-caps). The inner most part of the detector is the tracking system, designed to reconstruct the tracks and vertices of charged particles. The strong 2 T magnetic field created by a surrounding solenoid magnet bends the trajectories of charged particles and allow for momentum measurements with high resolution.

Beyond the tracking is the high granularity calorimetry system, which measures the energy of individual electrons photons and jets. The muon spectrometer is the outer most part of the detector as the muon is the only particle other than the neutrino to completely traverse the detector. There are ten toroidal-shaped magnets embedded in the muon spectrometer thereby allowing further momentum measurement and distinction of muons from anti-mouns. A more detailed description of the subdetectors is given in the following sections.

It is now convenient to define variables and the coordinate system that is used in this analysis. ATLAS uses a right-handed coordinate system where the z direction is defined in the direction along the beam line and the y direction points vertically upwards from the centre of the detector. The azimuthal angle ϕ is the angle in the transverse plane around the beam line and the polar angle, θ, is the angle from the beam line. The momentum and energy in the transverse plane are defined as

$$p_T = \sqrt{p_x^2 + p_y^2} \tag{2.2}$$

and

$$E_T = E \sin\theta \tag{2.3}$$

The p_T and E_T variables are used because these are invariant under Lorentz transformations in the direction of the beamline. The rapidity

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z} \tag{2.4}$$

is also invariant under Lorentz transformations and for high energies where $p \gg m$, y approximates

to the pseudorapidity, η, which can be expressed in terms of the angle θ.

$$\eta = -\ln(\tan\frac{\theta}{2}) \tag{2.5}$$

For particles which do not interact with any part of the detector, such as a neutrino, the energy is not measured and is interpreted 'missing' energy $E_{T,miss}$.

Separation of two particles a and b, $\Delta R_{a,b}$ is measured in $\eta - \phi$ space using

$$\Delta R_{a,b} = \sqrt{(\eta_a - \eta_b)^2 + (\phi_a - \phi_b)^2}. \tag{2.6}$$

The purpose of the ATLAS experiment is to obtain precise measurements of physical phenomena and to search for new physics beyond the SM. One of the main objectives of the experiment is to determine the nature of spontaneous symmetry breaking through the discovery of the Higgs boson. Now that a Higgs boson has been discovered measurements are being made to determine the cross sections, branching ratios, spin, mass and its couplings. The Higgs boson, isn't detected directly but through its decay products, which can be detected with the ATLAS detector with great precision.

2.2.1 Trigger and Data Acquisition

Most events in pp collisions are low energy scattering processes which are not interesting to study. The interesting events come from high E_T hard scatters or events with high $E_{T,miss}$. Keeping every event would not be feasible given that there were bunch crossings every 50 ns in the data collected so far. A trigger system is therefore in place to only save the events that are of interest.

The ATLAS trigger system operates at three levels: L1, L2 and the event filter (EF). L1 reduces the rate from 20 MHz rate of data taking to 75 kHz. It is required to be fast as new events are occurring every 50 ns. Regions of interest (RoI) in η and ϕ space from slices of the detector are identified based on reduced granularity information from the calorimetry system and the muon spectrometer. A decision is made whether or not to keep the event by the central trigger processor, based on energy thresholds and other interesting event characteristics.

After an event is accepted by L1, the event information is stored in readout buffers, whilst a decision is being made by the L2. Full access of all the information within the RoI from all the

subdetectors is available to the L2 trigger. The existing information is made more precise with fast algorithms to reconstruct particle tracks and other features of the event.

If the event is accepted by L2, full event reconstruction takes place using the same algorithms that are used for the offline analysis after the data is stored. These provide better threshold measurements and particle identification. After the EF the data taking rate is reduced to 400 Hz and the data are stored for offline analysis.

There are different triggers for different types of physics processes that are of interest to store. If a trigger is occuring at a high rate it can be 'prescaled' meaning the acceptance is reduced by a 'prescale' factor.

During data taking each run is divided into luminosity blocks. This is so the prescales can be changed as the luminosity progressively decreases during a run. In case of large dead time or part of the detector is not responding the corresponding lumi blocks can be rejected, whilst leaving the integrity of the rest of the run intact [29].

2.2.2 The Inner Detector

The inner detector is the tracking system which consists of three components that cover a pseudo-rapidity up to $|\eta| < 2.5$, as shown in Figure 2.2. As particle tracks will be much closer together nearest to the interaction point, the tracker is designed to increase in granularity with decreasing radius. The purpose of the inner detector is to reconstruct the tracks of charged particles. This relies on either semiconductor detectors or gas ionisation.

Pixel Detectors

The inner most part of the tracking system uses pixel semiconducting technology to provide high resolution track reconstruction. In the barrel, there are three concentric layers of pixel modules and three disks in each of the end-caps, which accurately determine three space points of the particle tracks. It is positioned closest to the interaction point, extending 650 mm in z and 122.5 mm in detector radius, r. Being so close to the beam, each module has to be radiation hard. The overall resolution of the pixels can reconstruct tracks which are 100 μm in z and 15 μm in $r\phi$ [31, 32]. Not only does this provide accurate primary vertexing but also the ability to distinguish multiple vertices apart in high pile-up events, and measuring displaced secondary vertices from long lived

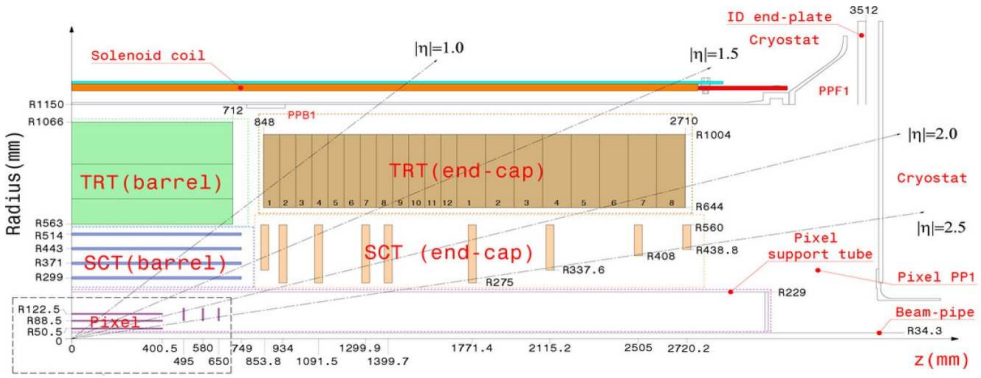


Figure 2.2: Detailed layout of all the components of the ATLAS inner detector in the $y-z$ plane including the pixels, semiconductor tracker and the transition radiation tracker. Absolute pseudorapidity is marked every $|\eta|$ =0.5 up to 2.5, the maximum tracking coverage, $|\eta| = 2.5$ [29].

particle decays, as shown in Figure 2.3.

Semiconductor Tracker

The semiconductor tracker (SCT) uses similar semiconducting technology as the pixel detector but with silicon microstrips. Each SCT module has two sensors with longitudinal strips, glued back-to-back with one at a 40 mrad stereo angle to provide hit measurement in $z-\phi$. The SCT has reduced granularity relative to the pixel detector, however the occupancy is much lower as the SCT is positioned further away from the interaction points. The SCT can achieve a resolution of 17 μm perpendicular to the strips and 580 μm parallel to the strips [33].

The modules are arranged in four concentric layers in the barrel region designed to provide four space point position measurements of charged tracks. In each of the end-caps, the modules are arranged radially in nine layers.

Transition Radiation Tracker

The TRT is the outer most part of the ATLAS tracking system and is the largest. It is a gaseous detector comprised of many straws 4 *mm* in diameter. These are arranged parallel to the beam axis in the barrel region and spoking out radially in the end-cap region. With little material the chance of photon conversion is minimised and the many straws can be used to precisely measure ≈ 30 positions along a particle track.

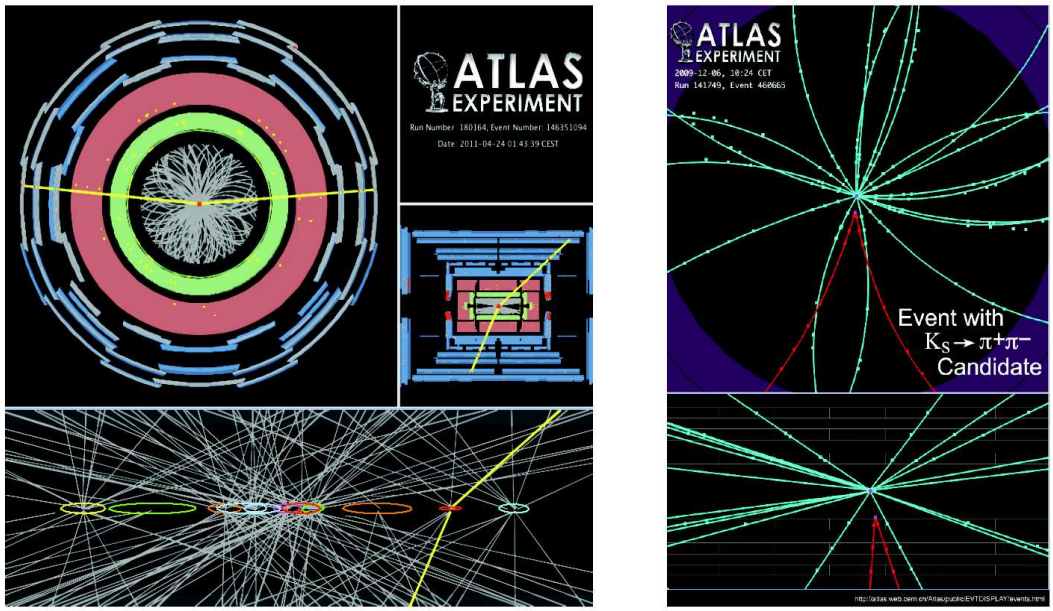


Figure 2.3: A display of multiple interactions in a single bunch crossing from ATLAS data. Eleven pp vertices have been identified (left). Amongst all this activity a secondary vertex, likely to be coming from a K_s particle (left), has also been indentified [31].

The gas mixture in the straws consists of Xe, CO_2 and O_2 and a gold plated wire runs through the centre. The wire is grounded and a negative voltage is applied to the straw. When a charged particle traverses the straw the gas molecules are ionised to electron-ion pairs. The electrons drift towards the wire and positive ions drift towards the straw. As the electrons accelerate, they gain enough momentum to produce more electron-ion pairs causing an electron avalanche. The build up of charge on the wire produces a voltage pulse, which is interpreted as the signal of a charged particle crossing the straw.

The gas has a high concentration of Xe, chosen for the high absorption efficiency of transition radiation (TR). TR is produced when charged particles with a high Lorentz factor pass through materials of different dielectric constants. Such radiating material occupies the space between the straws. This is useful to discriminate electrons from pions, as electrons produce more transition radiation than pions [29].

2.2.3 The Electromagnetic Calorimeter

The electromagnetic calorimeter (ECAL) provides a detector coverage of electromagnetically interacting particles of $|\eta| < 1.475$ in the barrel and between $1.375 < |\eta| < 3.2$ in the end-caps. It

is composed of modules as seen in Figure 2.4, designed with accordion shaped layers of absorbers and copper-kapton electrodes immersed in liquid argon. This accordion-shaped geometry gives full azimuthal coverage. Detection is achieved by the initiation of an electromagnetic shower in the absorbers which then ionises the liquid argon. Electrons drift towards the electrodes and a pulse is read out in cells of $\Delta\phi$ and $\Delta\eta$, which are then converted into an energy measurement. The full depth of the ECAL corresponds to over 20 radiation lengths, so that electrons and photons are fully absorbed before reaching the hadronic calorimeter.

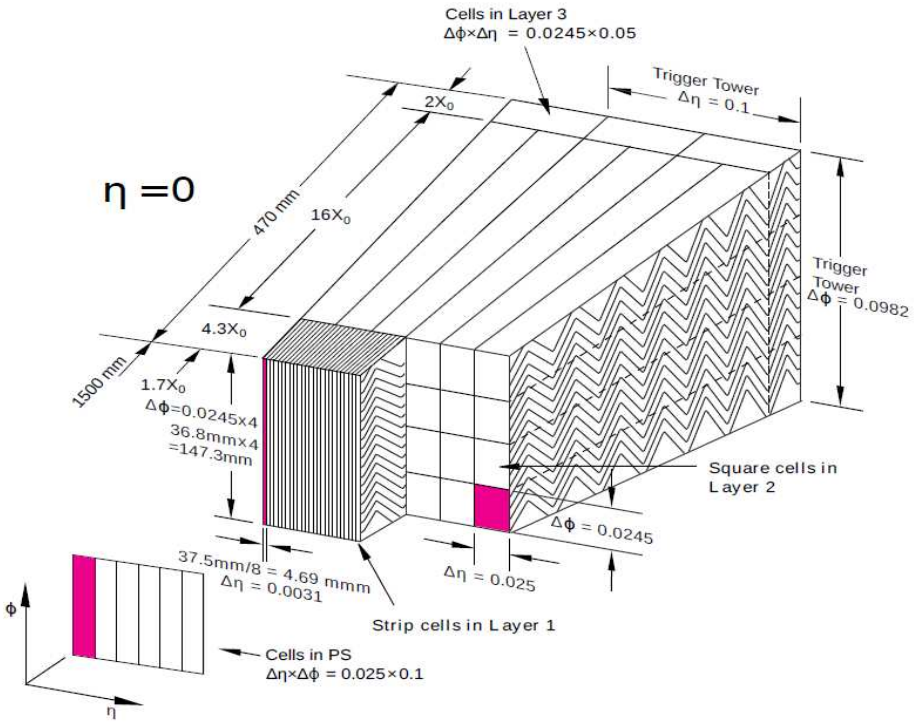


Figure 2.4: Schematic showing the dimensions of the 3 samplings and the presampler in an ECal module at $\eta = 0$ [34].

Each module is divided into three sampling layers, with different size cells in each layer. The first sampling has the finest granularity that is made up of strip cells that are more granular in η ($\Delta\eta \times \Delta\phi = 0.0031 \times 0.098$). An important aspect of this design is that it can provide a positioning measurements for photons, which can not be achieved by the tracker. It can also discriminate between real photons and π^0 hadrons which decay into two photons separated by small ΔR. This is demonstrated in Figure 2.5 which shows the energy deposits in the first and second samplings for

a photon and a π^0. In the second sampling (where the majority of the energy is deposited) the cells are less fine and the two candidates appear to be similar. However in the first sampling the energy from the π^0 candidate is detected in two 'clusters' because of the fine granularity, suggesting that there are two photons which are very close together. The outer sampling has less fine granularity, and is used for triggering purposes.

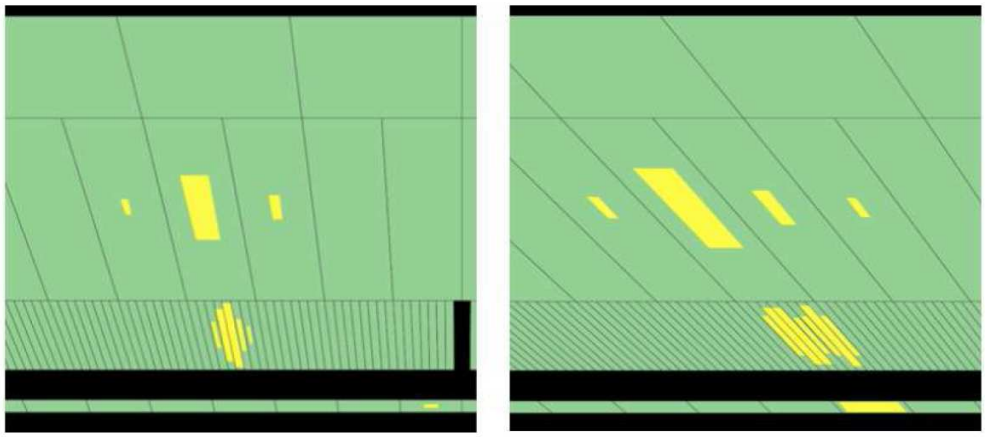


Figure 2.5: Detection of electromagnetic energy in the first and second samplings of the ECAL. A photon candidate is shown on the left and a $\pi^0 \to \gamma\gamma$ candidate is shown on the right [34].

Due to energy losses of electrons and photons upstream from the ECAL, there is a pre-sampler in place before the first sampling of the main ECAL module to provide an estimate of the energy losses.

2.2.4 The Hadronic Calorimeter

The hadronic calorimeter (HCAL) provides detector coverage of hadronically interacting particles up to $|\eta| < 4.9$. Full coverage is useful for analyses which require a measurement of missing energy. The absorbing material is required to be dense to ensure that all particles other than muons are absorbed before reaching the muon spectrometer ($|\eta| = 0$ corresponds to 9.7 interaction lengths).

The barrel region consists of scintillating tiles alternating with steel plates. The scintillating light is measured by photomultiplier tubes. The tiles are oriented radially and at a normal to the beam line for full azimuthal coverage [29].

The hadronic calorimetry in the end-caps also uses liquid argon detection. These are the

hadronic end-cap calorimeters (HEC) covering $1.5 < |\eta| < 3.2$ and the forward calorimeters (FCAL) covering $3.1 < |\eta| < 4.9$, which is important for the measurement of forward jets. Each HEC consists of wheels that are made of wedge-shaped modules of flat copper plates oriented at a normal to the beam line. The readout provides granularity of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ for $|\eta| < 2.5$ and $\Delta\eta \times \Delta\phi = 0.2 \times 0.2$ otherwise [29].

The FCAL is divided into three components. To reduce the neutron albedo back into the inner detector the FCAL is positioned further out from the interaction point relative to the HEC. As a result the material is required to be more dense to ensure full absorption. The inner most FCAL module uses copper absorbers for electromagnetic measurements and tungsten absorbers for the two outer most components [35].

2.2.5 Muon Chambers

The muon spectrometer is positioned on the outer most part of the detector. It is divided into three layers to give precision coverage within $|\eta| < 2.7$. Each layer is made up of specific components that are arranged concentrically in the barrel and consist of disks in the end-caps. A magnetic field is provided within the layers by 10 toroidal magnets (eight in the barrel and one in either end-cap).

The muons trajectories will bend in the toroidal magnetic field and the amount of bending is measured, to determine the momentum of the muons. The momentum resolution ranges between 4% for muons of 3 GeV and 10% for a muon of 1 TeV [36].

The measuring components consist of drift tubes, cathode strip chambers, and two types of triggering devices: the resistive-plate chambers and the thin-gap chambers. The drift tubes work on the principle of charged muons ionising the gas within. Electrons are liberated and drift towards a wire due to the applied electric field. This robust design has the advantage that the wires produce a radially symmetric electric field, so the drift time has little dependence on the muon incidence angle.

All the precision measurements are provided by the drift tubes except in the inner most layer of the end-caps with $|\eta| > 2$, where the drift tubes are replaced by cathode strips, due to the high occupancy rate in the forward region and their better resolution in the bending plane [29].

Chapter 3

Signal and Background Processes

The data analysis presented in this thesis is concerned with the $H \to \gamma\gamma$ decay channel. In this chapter the relevant signal topologies will be described, and the corresponding SM backgrounds will be discussed. A description of Monte Carlo (MC) event simulation is given and the use of different MC generators is discussed to model the events. An argument is put forward to use $13\,\text{fb}^{-1}$ of data to model the background. A signal region in the invariant mass window, $120 < m_{\gamma\gamma} < 130\,\text{GeV}$ is defined to contain the majority of the signal events. Outside the signal region, the real data is expected to consist almost entirely of background.

3.1 Signal Processes

Although the $H \to \gamma\gamma$ branching fraction is extremely small (0.228% for $m_H = 125\,\text{GeV}$ [21]) this decay is one of the best channels to detect and study a light Higgs boson due to its clean signature of two isolated high p_T photons and the excellent experimental mass resolution. The invariant mass between the two leading photons γ_1 and γ_2, defined as:

$$m_{\gamma\gamma} = \sqrt{2E_{\gamma 1}E_{\gamma 2}(1-\cos(\theta_{\gamma 1\gamma 2}))} \tag{3.1}$$

where $E_{\gamma 1}$ and $E_{\gamma 2}$ are the respective energies of the two leading photons and $\theta_{\gamma 1\gamma 2}$ is the opening angle between the two photons. Since the photons are well defined objects and can be measured with very good energy resolution, the signal events cluster in a peak in a narrow mass window, as shown in Figure 3.1.

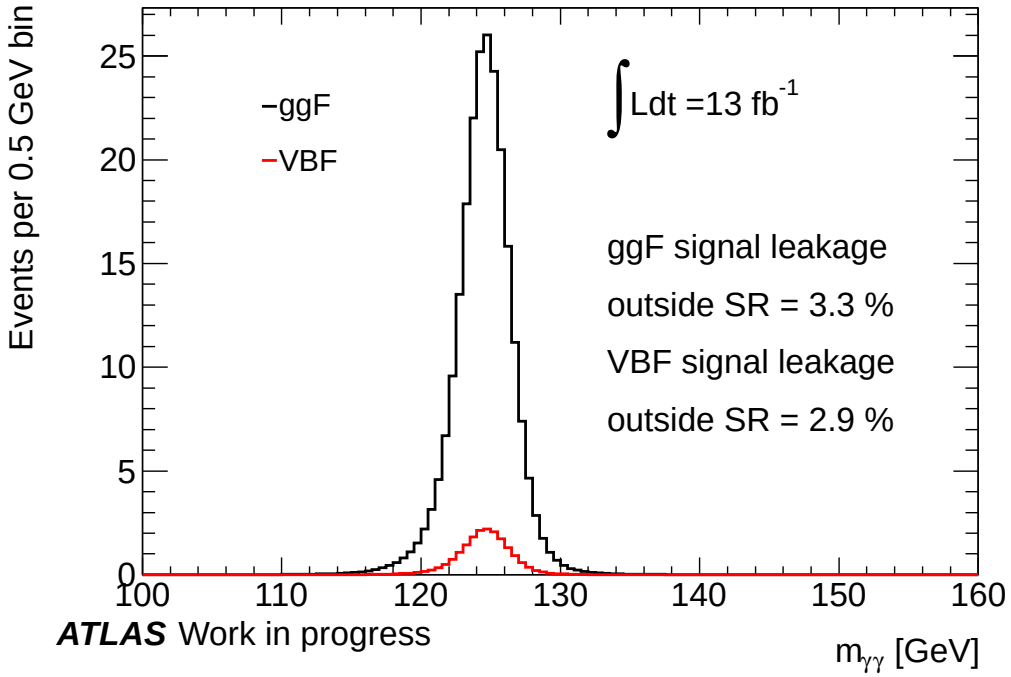


Figure 3.1: $m_{\gamma\gamma}$ calculated for MC $H \rightarrow \gamma\gamma$ signal events containing two high p_T photons that have been simulated using Pythia and Powheg for both the gluon-gluon fusion and the VBF mechanisms. The gluon-gluon fusion signal distribution is shown in black with the VBF signal distribution (in red) superimposed. The distributions are normalised to 13 fb^{-1}.

At the LHC, the Higgs boson is produced by five mechanisms, which possess certain features such as additional jets or leptons that are tagged to provide extra signal sensitivity. In particular two additional 'tag jets' are present in Higgs boson events where the Higgs boson is produced by the VBF mechanism. The tag jets are formed from the quarks in the incoming protons fragmenting after recoiling from the weak bosons that fuse to produce the Higgs boson. Jets may also be produced in gluon-gluon fusion signal events but these jets are initiated from higher order QCD radiation, and are mistaken as tag jets. The tag jets in VBF signal events are much more forward and have higher p_T when compared with those of the gluon-gluon fusion events, shown by a much higher multiplicity of jets in VBF at higher pseudorapidity values (see Figure 3.2). The tag jets in VBF signal are often detected in opposite ends of the detector and are separated by a large gap in pseudorapidity space, $\Delta\eta_{jj}$. The dijet system of the two tag jets usually has a large invariant mass M_{jj} and the azimuthal angles of the dijet system in the transverse plane, ϕ_{jj}, and the diphoton system, $\phi_{\gamma\gamma}$, are expected to be separated by approximately 180°, for momentum conservation

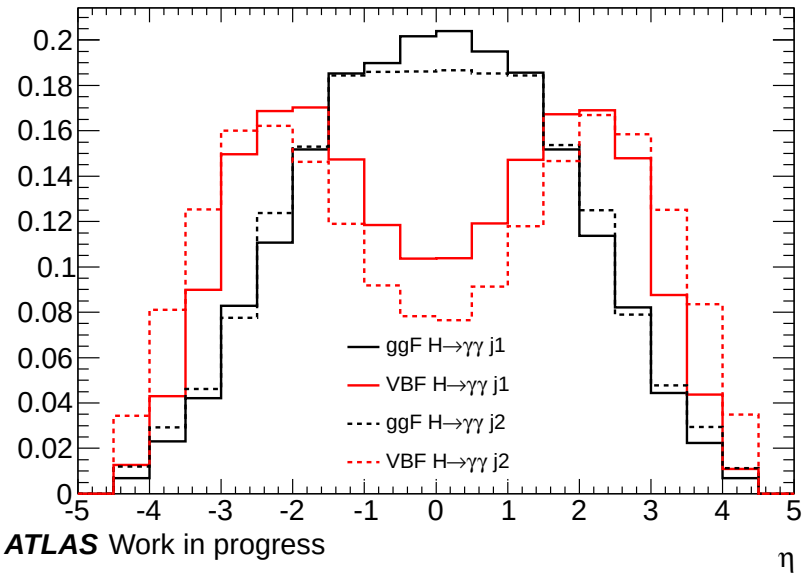


Figure 3.2: Pseudorapidity of the leading (j_1) and subleading jets (j_2) for events which contain at least two photon candidates and at least two jet candidates. Comparison between simulated signal samples that have been generated with Pythia and Powheg for both for gluon-gluon fusion and VBF.

reasons. This property is quantified by the variable:

$$\Delta\phi_{jj,\gamma\gamma} = |\phi_{jj} - \phi_{\gamma\gamma}| \tag{3.2}$$

This is also reflected in the balance between the transverse momenta of the diphoton system $\vec{p}_{T,\gamma\gamma}$ and the dijet system $\vec{p}_{T,jj}$. The p_T balance variable will now be defined as $|\vec{p}_{T,\gamma\gamma} + \vec{p}_{T,jj}|$ which is zero if the jets and photons are perfectly balanced. The two jets and two photons are more likely to be in balance in VBF events as shown in Figure 3.3.

As a large separation between the jets and the decay products is also expected, a separation variable between the leading photon and leading jet is defined as:

$$\Delta R_{\gamma 1,j1} = \sqrt{(\eta_{\gamma 1} - \eta_{j1})^2 + (\phi_{\gamma 1} - \phi_{j1})^2} \tag{3.3}$$

In associated production events with a weak boson, the weak boson can decay leptonically or hadronically. Where the weak boson decays leptonically, an electron or a muon can be observed in addition to the two photons. Where the weak boson decays hadronically, there are two jets observed and the reconstructed M_{jj} is therefore similar to the mass of the W or Z boson, as shown

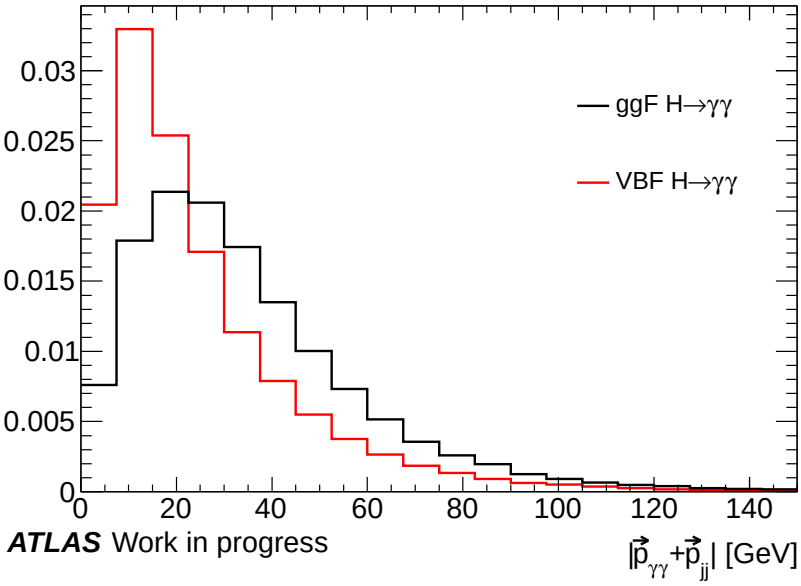


Figure 3.3: Comparison between Monte Carlo signal samples for gluon-gluon fusion and VBF of the p_T balance calculated for every event containing at least two photon candidates and at least two jet candidates.

in Figure 3.4

The $p_{T,t\gamma\gamma}$ variable [37] is the magnitude of the vector sum of the p_T of the two leading photons projected onto a trust axis $\hat{t}$.

$$p_{T,t\gamma\gamma} = |(\vec{p}_{T,\gamma 1} + \vec{p}_{T,\gamma 2}) \wedge \hat{t}| \tag{3.4}$$

where

$$\hat{t} = \frac{\vec{p}_{T,\gamma 1} - \vec{p}_{T,\gamma 2}}{|\vec{p}_{T,\gamma 1} - \vec{p}_{T,\gamma 2}|} \tag{3.5}$$

$p_{T,t\gamma\gamma}$ is generally high in associated Higgs boson production with W or Z bosons. and the jets are separated by a small pseudorapidity gap.

Associated production with $t\bar{t}$ is the least likely Higgs production mechanism which leaves a signature of two photons and multiple jets due to the decay chains of the top quarks.

3.2 Background Processes

There are various types of background to the $H \rightarrow \gamma\gamma$ signature, which can be both irreducible and reducible. The irreducible background is other pp events, which include final state isolated photons. The main diphoton backgrounds are shown in Figure 3.5. The Born process is where

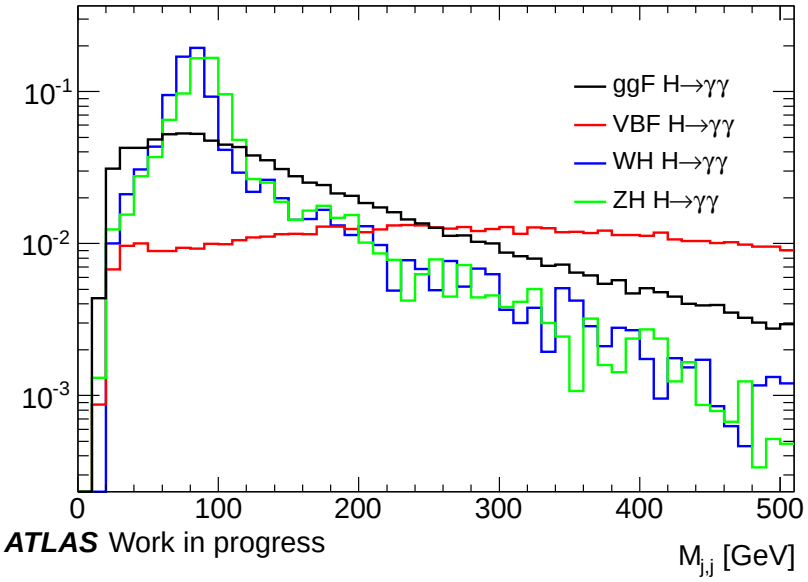
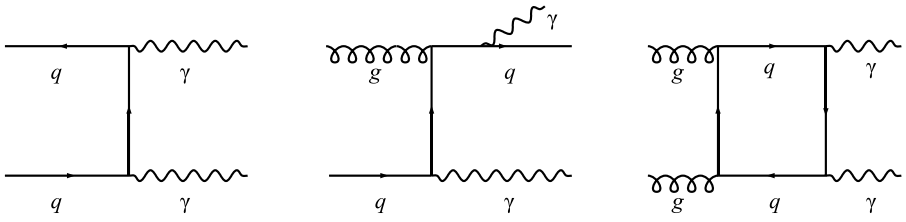


Figure 3.4: Comparison between simulated signal samples for gluon-gluon fusion and VBF of the invariant mass of the two leading p_T jets for every event containing at least two photon candidates and at least two jet candidates.

two photons are produced by two quarks as shown in Figure 3.5(a). A quark gluon interaction where two bremsstrahlung photons are radiated from the quarks, as shown in Figure 3.5(b) and a higher order box diagram, where two photons are produced from a gluon interaction, is shown in Figure 3.5(c).



(a) Born (b) Bremsstrahlung (c) Box

Figure 3.5: Leading order Feynman diagrams of the irreducible background processes for the $H \rightarrow \gamma\gamma$ signal.(a) $qq \rightarrow \gamma\gamma$, (b) $qg \rightarrow \gamma\gamma$ and (c) $gg \rightarrow \gamma\gamma$.

There are also many reducible photon backgrounds mostly dominated by leading neutral mesons in $\gamma-$jet or dijet events, (jets are plentiful in hadron colliders). However by using information from the inner detector and exploiting the information and fine granularity from the ECAL, the faking of photons by jets is reduced.

3.3 Signal and Background Modelling

3.3.1 Monte Carlo Simulation

The proton is a composite particle containing quarks and gluons (partons). Since the pp interactions are random, they are modelled by probability density functions (PDFs). For a process where parton i in proton a and parton j in proton b goes to particle k, the cross section is given by

$$\sigma_{ij\to k} = \int dx_a \int dx_b p_i^a(x_a, Q^2) p_j^b(x_b, Q^2) \hat{M}_{ij\to k} \tag{3.6}$$

where x_a is the fraction momentum of proton a, the PDF $p_i^a(x_a, Q^2)$ is the probability of parton i in proton a having x momentum at a momentum scale Q^2 and $\hat{M}_{ij\to k}$ is the amplitude (or matrix element) of the $ij \to k$ process. Events are simulated in accordance to the PDFs using Monte Carlo (MC) simulations.[38]

For hadron colliders, it is not just the hard scatter that the event generator must consider. Various other processes need to be taken into account when generating events, these are shown in Figure 3.6. Radiative corrections mean that incoming and outgoing particles may radiate out initial or final state photons or gluons. The radiated gluons would then hadronise producing extra jets in the event. The proper way to model radiative corrections would be to include them in the matrix element, except not all higher orders are known. Instead parton showering is used, which bases initial and final state radiation on parameters determined from data.

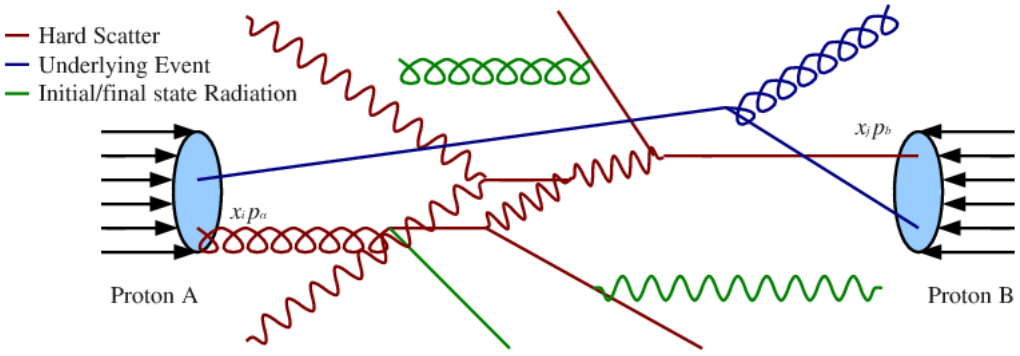


Figure 3.6: Schematic diagram of the typical scattering and radiative processes in pp collisions.

Since quarks and gluons carry colour charge, they have to be confined and hadronise at distance

scales O (1) fm. Heavy and offshell hadrons decay in to other hadrons until stable states are reached producing jets in the event.

In addition to the hard scatter, there are multiple soft scattering processes, this is known as the underlying event [39] which can also produce associated initial and final state radiation. QCD radiation hadronises and produces even more jets in the event.

The signal in $H \to \gamma\gamma$ is modelled by either Pythia [38, 40] alone or Pythia with Powheg [41, 42]. The associated production mechanisms are modelled using just Pythia, where recent updates take into account updated information from the particle data group [19] and more accurately simulate the transverse momentum of the Higgs boson. The gluon-gluon fusion and VBF processes are the highest cross section processes and have large higher order corrections. Accurate modelling of the jets is of additional importance in the VBF signal, as cuts are applied to the tag jets. These processes are modelled with Powheg, which can account for the higher order corrections, including QCD and electroweak corrections. Next-to-leading order (NLO) inputs from Powheg are interfaced with Pythia, and Pythia is used to model the hadronisation and parton showering of the gluon-gluon fusion and VBF processes.

Due to high luminosity in the 2012 data, there were several pp collisions in every bunch crossing. In order to simulate high pileup multiple minimum-bias proton scattering events are generated and overlaid on the hard scatter.

In this analysis a total of 15 MC samples were used to simulate the $H \to \gamma\gamma$ events produced by gluon-gluon fusion, vector boson fusion and the associate production processes. Each process was simulated at three different Higgs boson masses (m_H). More events were generated for the gluon-gluon fusion process and the VBF process at 125 GeV. As these two processes are measurable at 13 fb^{-1}, high statistics are advantageous. The relevant information for each simulated sample is shown in Table 3.1.

The cross sections used to normalise the number of events are shown in Table 3.1. These are calculated with much higher precision compared with the samples used. Gluon-gluon fusion is calculated at next-to-next leading order (NNLO) and next-to-next leading logarithmic (NNLL) order for QCD processes and NLO for electroweak processes. VBF, WH and ZH is calculated at NNLL for QCD processes and NLO for electroweak processes and ttH is calculated at NLO for QCD processes [21].

Geant 4 [43] was used to simulate the detector effects and the interactions of the final state

particles, passing through the detector material.

3.3.2 Data-Driven Approach to Background Estimation

The background events have to be modelled accurately in order to extract the H$\rightarrow \gamma\gamma$ signal. It is possible for this to be done with the MC simulations but as mentioned in Section 3.2, the MC simulation is limited by not knowing all higher order corrections. There are also various composite background processes, which would be difficult to determine. Studies of VBF events involve cuts on jets, so any uncertainty in the parton showering and hadronisation propagates into the systematic uncertainty in the γ+Njets samples. An alternative way to model the background is to use data events to model it, which is what has been done for this analysis.

Data can be used to model the background providing there is no contamination from the signal. Recent measurements on the Higgs mass conclude $m_H = 125.5 \pm 0.2\text{stat} \pm^{0.5}_{0.6} \text{syst}\,\text{GeV}$ [10] and it was shown in Figure 3.1 that for a simulated $m_H = 125\,\text{GeV}$, $m_{\gamma\gamma}$ occupies a narrow mass window. A signal region is therefore defined as the invariant mass window $120 < m_{\gamma\gamma} < 130\,\text{GeV}$. Assuming the actual Higgs mass is close to 125 GeV it is very unlikely that genuine Higgs events will have $m_{\gamma\gamma} < 120\,\text{GeV}$ or $m_{\gamma\gamma} > 130\,\text{GeV}$, therefore it is a very good approximation to assume that all events in the lower sideband region $100 < m_{\gamma\gamma} < 120\,\text{GeV}$ and the upper sideband region $130 < m_{\gamma\gamma} < 160\,\text{GeV}$ are background events.

$H \to \gamma\gamma$ Production	m_H [GeV]	σ [pb]	Branching ratio [%]	Expected Events at 13 fb^{-1}	N Events
ggF	120	21.13	0.233	64.0	299999
ggF	125	19.52	0.228	57.9	2984986
ggF	130	18.07	0.225	52.9	99997
VBF	120	1.649	0.233	4.99	100000
VBF	125	1.578	0.228	4.68	979993
VBF	130	1.511	0.225	4.42	49999
WH	120	0.7966	0.233	2.41	30000
WH	125	0.6966	0.228	2.06	30000
WH	130	0.6095	0.225	1.78	29900
ZH	120	0.4483	0.233	1.36	29997
ZH	125	0.3943	0.228	1.17	30000
ZH	130	0.3473	0.225	1.02	30000
ttH	120	0.147	0.233	0.445	30000
ttH	125	0.1302	0.228	0.386	30000
ttH	130	0.1157	0.225	0.338	29999

Table 3.1: Assorted statistics for 15 MC $H \to \gamma\gamma$ signal samples used in this analysis for 5 different processes generated for 3 different values of m_H. Cross sections and branching ratios are obtained from Reference [21]. Gluon-gluon fusion is calculated at NNLO+NNLL QCD + NLO EW. VBF, WH and ZH is calculated at NNLL QCD + NLO EW and ttH are calculated at NLO QCD.

Chapter 4

Reconstruction of Physics Objects

This analysis requires the identification of various particles and physics phenomena. As this analysis is concerned with the Higgs boson which decays via the diphoton channel, efficient identification of photons is required. In addition, the various production mechanisms are studied (the VBF and associate production with vector bosons) which require identification of jets, electrons and muons. The reconstruction and identification of these physics objects is described in this chapter.

4.1 Photons

Photons are electromagnetically interacting particles and have no electric charge. They are therefore identified by the presence of an electromagnetic cluster in the ECAL with no associated track. This is only true however, for photons which do not convert into electron-positron pairs. As the photons interact with material in the detector, as many as 60% [44] convert before they reach the ECAL. During LHC run time, photons which have converted are initially classified as electrons and are then later recovered in the offline analysis. The electrons which are considered converted photon candidates, are those which have a conversion vertex associated with the track, or are associated with tracks that are not consistent with tracks made by prompt electrons. A converted photon, is recovered providing it can be matched with energy clusters in the second sampling in the ECAL, within an η-ϕ window. Tracks and vertices are then refitted under the electron hypothesis in order to correct for bremsstrahlung energy losses.

Hadronic background such as $\pi^0 \rightarrow \gamma\gamma$ are distinguished from other photons by utilising information from the calorimetry systems and by applying isolation cuts around the photon candidate.

Using the calorimetry information, so called shower shape variables are defined, where one can apply cuts to discriminate between prompt photons coming from the hard interaction and $\pi^0 \rightarrow \gamma\gamma$. A study of MC samples [45] identified these variables and compared each one for 'real' photons and 'fake' photons; this is shown in Figure 4.1. Real photons were defined here as photons that are reconstructed from events in γ-jet MC samples and can be matched up with a true (MC) photon from the hard scatter. The fake photons are those reconstructed from dijet MC samples that are not matched with true photons from parton bremsstrahlung.

The shower shape variables describe three key distinguishing features between jets and photons [45, 46]: hadronic leakage, lateral showering and substructures in the showers. The hadronic leakage measures the ratio R_{had} of transverse energy deposited in the first sampling of the HCAL and in the cluster in the ECAL. Real photons are electromagnetically interacting particles, so R_{had} has a low value, whereas jets contain hadronic particles and initiate hadronic showering in the HCAL, therefore R_{had} has a high value. Lateral showering is measured because photons produce narrow clusters, whereas the jets are more broad. This is measured using the following shower shapes:

- R_η is the ratio of ECAL energy in a 3×7 ($\Delta\eta \times \Delta\phi$) group of cells and the energy in a 7×7 group of cells;
- R_ϕ is the ratio of ECAL energy in a 3×3 ($\Delta\eta \times \Delta\phi$) group of cells and the energy in a 3×7 group of cells;
- w_2 is the ECAL shower width in η in a window of 3 cells, using the energy weighted sum of all cells.

The substructure of the showers is measured using the ultra-fine strip layers in the 1^{st} sampling of the ECAL. This is to distinguish real photons from neutral mesons that have decayed to two photons that are close together. Without the fine granularity, this would appear to be one photon but with the strip layers, it is possible to resolve two energy maxima (E_1 and E_2; $E_1 > E_2$) in the ECAL cluster and an energy minimum E_{min} in between. The following shower shape variables are used:

- $\Delta E = E_2 - E_{min}$;
- $E_R = (E_1 - E_2)/(E_1 + E_2)$;

- The fraction of total energy that is deposited outside of the 3 strips centred on the cluster;
- w_3 the width of the cluster over 3 strips around one of the maximum energy deposites weighted by the measured energy in each strip;
- w_{tot} the width of the cluster over the number of strips that have the same η as 2.5 cells in the second layer.

For 2012 data, the cuts on the shower shape variables have been optimised for high pileup conditions [9].

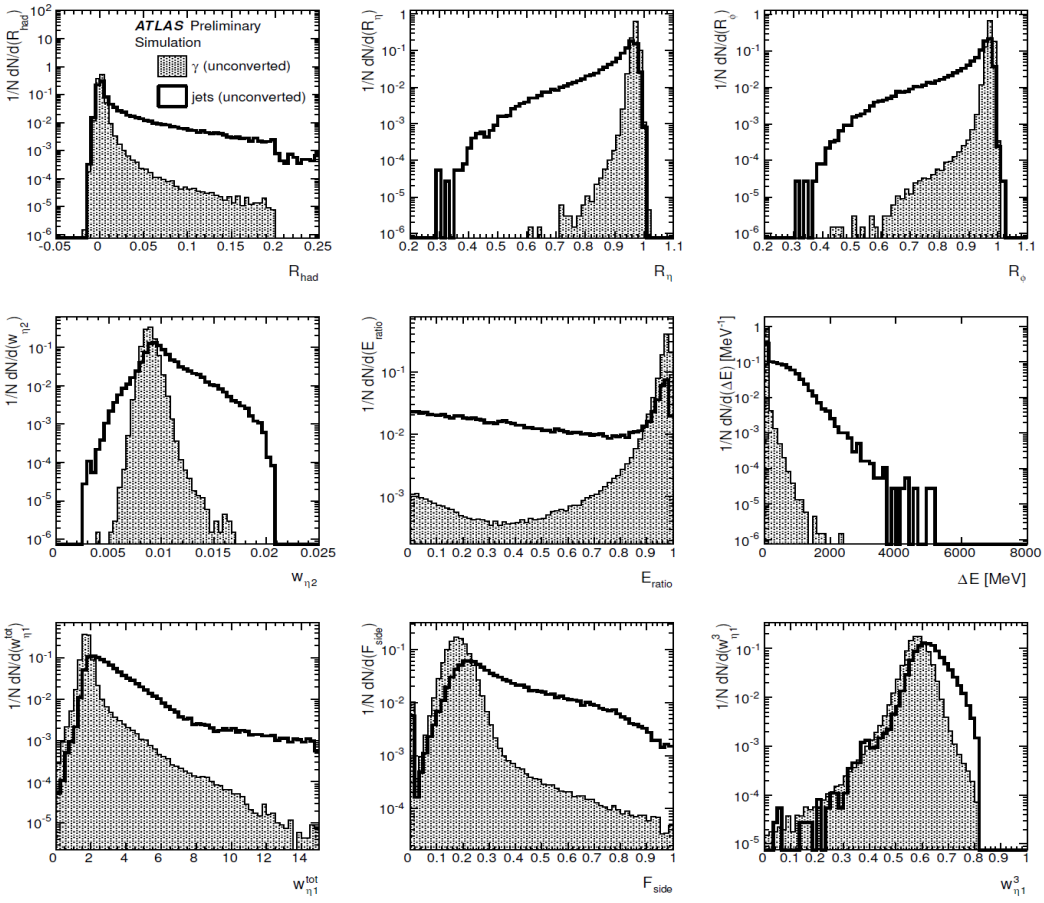


Figure 4.1: Shower shape variables for unconverted real and fake photons of $E_T > 20\,\text{GeV}$, as described in [45]. For each variable distribution Distributions are normalised for shape comparison.

The quality of the photon candidate is either loose or tight. “Loose” requires the photon candidate passes cuts based on R_{had}, R_η and w_2. Loose quality cuts are sufficient enough to identify

photons at trigger level but for offline physics analysis the photon candidates are required to satisfy "tight" cuts, which implement all of the shower shape variables. In addition an acceptance cut is applied, $|\eta| < 2.37$, since outside the tracking range photons and electrons become indistinguishable.

As the selection of two photons is required in the $H \rightarrow \gamma\gamma$ analysis, it is useful to first determine the common primary vertex (PV) containing the hard scatter from which the two photons would have in principle originated. The pseudorapidity measurements of the two photons are corrected to that of a pseudorapidity coming from the chosen PV. The measured p_T is then corrected using the corrected pseudorapidity. This provides significant improvement to the signal mass resolution. Due to multiple pp interactions there are many PVs in a bunch crossing, so an artificial neural network (NN), multivariate analysis classifier is used to select the most likely candidate. For each collision vertex the inputs to the NN are:

- The sum of the squared p_T of tracks consistant with the vertex, $\sum p_{T,track}^2$;
- The scalar sum of track p_T consistant with the vertex, $\sum |\vec{p}_{T,track}|$;
- The difference in azimuthal angle between the diphoton system and the p_T vector sum of tracks consistant with the vertex;
- $z_{PV} - z_{point}/\sigma_{point}$ where z_{PV} is the position in z of the PV. For the case of unconverted photons, z_{point} is the z position extrapolated from pointing backwards from the clustering positions in different layers of the ECAL and for the case of converted photons, extrapolating from track positions measured by the SCT. σ_{point} is the resolution of the pointing (15mm for the unconverted photons and 6mm for the converted photons) [47].

As π^0 and other neutral mesons are usually accompanied by additional hadronic activity, further background suppression can be gained by applying transverse energy isolation (E_T^{iso}) cuts, determined from the calorimetry system, and transverse momenta isolation (p_T^{iso}) cuts, determined from the tracking system. E_T^{iso} is determined using the methodology described in [48]. The energy in a $\Delta R = 0.4$ cone in η and ϕ space around the photon is determined. On average, the majority of the photon energy is contained within the 5×7 cell region in the centre of the cone which is subtracted from the total energy in the cone. An E_T-dependent correction is applied to account for energy leakage outside the 5×7 region. A further correction is applied, to account for the ambient

energy contribution in the cone coming from the underlying event and pileup. This is determined by the average energy density in each event of all reconstructed jets using the k_T algorithm. If the remaining $E_T^{iso} > 6\,\text{GeV}$ it is likely that the photon is associated with hadronic activity and the photon is not used.

$p_{T,track}^{iso}$ is determined by constructing a $\Delta R = 0.2$ cone in η and ϕ space around the photon. The sum of all the track $|\vec{p}_T|$ is calculated and if this exceeds 2.6 GeV the photon is not used [9]. $p_{T,track}^{iso}$ is calculated using only those tracks with $p_T > 1\,\text{GeV}$ that are associated with the chosen PV, excluding tracks originating from photon conversions.

The rejection rate of jets with $p_T > 25\,\text{GeV}$ is approximately 1/5000 (i.e. 1 in every 5000 jets is accepted as a photon) [45].

4.2 Jets

A jet is a spray of hadronic particles that have originated from the fragmentation (hadronisation) of a quark or gluon. As hadronic particles pass through the calorimeter, part of their energy is deposited in calorimeter cells. If the energy in a given calorimeter cell exceeds an energy threshold, an algorithm is initiated which clusters together the energy deposits. These "clusters" of cells are then combined together to form a jet. The clustering algorithm that is used in this analysis is the anti-k_T algorithm [49] with a distance parameter of 0.4.

Due to various QCD processes, multiple jets are reconstructed in most events and the number of reconstructed jets can vary depending on the clustering algorithm being used. The anti-k_T algorithm has an advantage over other clustering algorithms as it will combine low energy clusters with neighbouring high energy clusters, before the low energy clusters can combine amongst themselves.

There are various forms of noise in the detector that can be wrongly reconstructed as jets, caused by:

- collisions between protons in the beams and gas molecules in the beam pipes;
- cosmic rays;
- calorimeter noise.

The noise from the calorimeter is the result of problematic cells in the HEC and ECAL. This noise is characterised by measuring the relative amounts of energy in each calorimeter cell and quantifying the quality and timing of the pulse shapes [50]. The suitability of the reconstructed jets for offline analysis is measured based on these characteristics.

The energy of the jets is initially determined by measuring the amount of electromagnetic energy deposited in the calorimeter, this does not take into account energy losses from dead calorimetry material, detector effects, energies of particles not measured by the calorimeter or particles which would, in truth, be part of the jet but were not reconstructed. The jets therefore have to be calibrated to their true energy. Before the calibration, two corrections are applied. The first correction takes into account the ambient energy contribution from pile-up, as a function of the number of primary vertices in the event and the pseudorapidity of the jet. The second corrects the pseudorapidity of the jet, assuming that it comes from the primary vertex of the hard scatter. The calibration is a correction of energy and direction, applied to each jet, as a function of its $E-\eta$. The corrections are derived from comparing jets in MC truth to jets in data for well understood kinematic processes [51].

In a MC study the "response" of the calorimeter to jets was measured after calibration in $\eta - p_T$ bins. The response is the ratio of the jet p_T compared to its matched truth jet. Any remaining deviations from unity in the jet p_T or energy response are used to calculate the systematic uncertainty for each $\eta - p_T$ bin in the signal samples [52]. This is further investigated in Chapter 8.

To suppress jets that are originating from pile-up interactions a jet vertex fraction (JVF) cut is applied to each jet [53]. The JVF is defined for each jet, as the ratio of the $|p_T|$ sum of the individual tracks, using only the tracks associated with the jet that originate from the chosen primary vertex, to the total scalar sum of all the tracks associated with the jet, irrespective of which primary vertex they originate from. Tracks coming from the primary vertices are associated with a jet, if ΔR between the reconstructed jet and the track is less than 0.4. A jet which originates from the hard scatter will have JVF≈ 1 and jets coming from the other pile-up vertices will have a JVF≈ 0. The JVF is set to -1 if the jet is outside the pseudorapidity region covered by the tracker.

Jets are used in this analysis if they pass loose quality requirements [50] and:

- $0.5 < |JVF| \leq 1$;
- the jet $p_T > 25\,\text{GeV}$ if $|\eta| < 2.5$ or $p_T > 30\,\text{GeV}$ if $|\eta| > 2.5$.

4.3 Electrons

The selection of electrons is similar to that of photons. Electrons are also electromagnetically interacting particles and therefore leave an energy deposit in the ECAL. However the electron is charged so the electromagnetic cluster is also required to be associated with a track. For electrons extra quality requirements are applied to remove ambiguity between electron and converted photons [54].

For this analysis the p_T of the clusters is required to be at least 15 GeV. The quality cuts applied are similar to the loose quality requirements for the photons, which are based on shower shape variables describing hadronic leakage and lateral showering profile. For electrons the pseudorapidity acceptance is increased to $|\eta| < 2.47$. In addition to the loose cuts, extra quality requirements are applied to the tracks based on hits in the inner detector tracking system, the extrapolation of the track to the cluster and the position of the track in relation to the chosen PV.

Cuts are applied on the transverse energy isolation, E_T^{iso}, from the calorimeter and transverse momentum isolation, p_T^{iso}, determined from the tracking system: $E_T^{iso} < 5\,\text{GeV}$ in a $\Delta R = 0.4$ cone around the electron and $p_T^{iso} < 3\,\text{GeV}$ in a $\Delta R = 0.2$ cone around the electron.

4.4 Muons

The muon is the only particle (other than neutrinos) not stopped by all of the material in the ATLAS detector. It is therefore reconstructed using the information from the outer most part of the detector, the muon spectrometer. The muon is also charged so information is also otained from the inner tracking system. For an object to be reconstructed as a muon several interactions are required in the pixels, SCT, and TRT.

Muon tracks are reconstructed in two ways. One method is to reconstruct the tracks using information from both the muon spectrometer and the inner tracker and combine the information together. If a track cannot be reconstructed properly in the muon spectrometer, the other way is to reconstruct a track in the inner detector and extrapolate to the muon spectrometer and determine if the track is associated with any interactions. The muons are required to have $p_T > 10\,\text{GeV}$ and be in a pseudorapidty region of $|\eta| < 2.7$.

To suppress background from cosmic rays, the minimum approach of the muon track to the

chosen PV is required to be no greater than 10 mm in the z direction and no greater than 1 mm in the transverse plane. In addition, the muon candidate has to be synchronised in time with the rest of the event. The muon also has an energy and track isolation $E_T^{iso} < 5\,\text{GeV}$ in a $\Delta R = 0.4$ cone around the muon, and $p_T^{iso} < 3\,\text{GeV}$ in a $\Delta R = 0.2$ cone around the muon.

4.5 Overlap and removing double counting

Double counting of physics objects can occur when a signal physical object is detected and it gets reconstructed as several different physics objects by independent reconstruction algorithms. If the separation, ΔR, between two types of physics objects is determined and is small, the two physics objects are said to be 'overlapped'. It is likely there is only one real physics object and the others are double counting. For example, see Figure 4.2, showing - for every event containing two tight isolated photons, the ΔR separation between the leading p_T photon and all the reconstructed jets. There is a very large number of jets very close to the photons. As the photons have passed tight isolated criteria, it is likely that these jets are duplicates of the photons and therefore these jets are removed from the event.

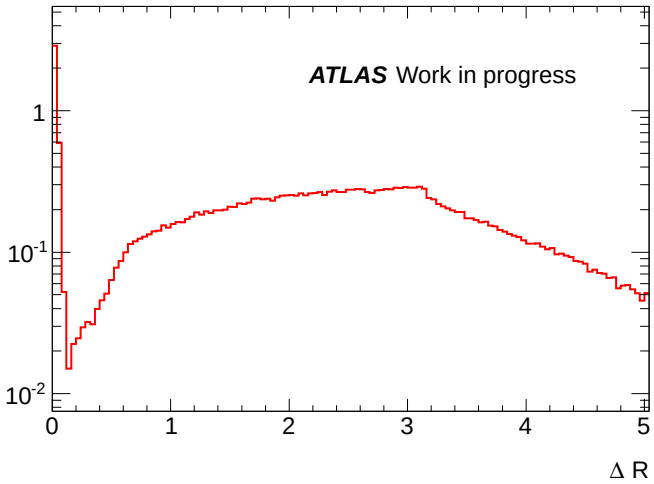


Figure 4.2: Histogram of ΔR between the leading p_T photon and all the jets in 10000 events from a VBF $H \rightarrow \gamma\gamma$ signal MC sample. The distribution is normalised to unity.

Chapter 5

Optimising the Selection of VBF $H \rightarrow \gamma\gamma$ Events

In July 2012 the ATLAS and CMS collaborations at CERN, announced the discovery of a new boson that had properties consistent with those of a Higgs boson as predicted by the SM [8, 55]. A measurement of the cross sections for the VBF and gluon-gluon fusion process will be another useful measurement to check for consistency with the SM.

In this chapter the procedure for selecting events with two high p_T photons is given. These events are categorised to separate the signals of the 5 production mechanisms. It will be shown using MC that the category intended to be enriched in the VBF signal needs reoptimising as the amount of VBF signal is limited and there is a noticeable amount of contamination from the gluon-gluon fusion signal. Two methods of reoptimisation have been investigated to increase the signal yield and the expected significance but at the same time reducing the amount of gluon-gluon fusion signal in the VBF enriched category. The two methods investigated were; A Boosted Decision Tree (BDT) classifier and changing the p_T thresholds defining the tag jets.

5.1 Event Selection of $H \rightarrow \gamma\gamma$ Events

The $H \rightarrow \gamma\gamma$ event selection follows the criteria set by ATLAS [47] at the end of 2012. Events must have fulfilled the requirement of a diphoton trigger that requires the presence of two electromagnetic clusters that have passed the loose photon quality criteria described in Section 4.1 and

transverse energy thresholds of $E_T > 35\,\text{GeV}$ for the most energetic and $E_T > 25\,\text{GeV}$ for the second most energetic , this trigger is at least 99% efficient at selecting those events which would pass the entire offline event selection. Each event selected must be from a lumi block and run that passed all data quality requirements. Events are removed from the analysis in the presence of noise in the calorimetry system. Events are selected if there is at least one primary vertex with at least three associated tracks.

5.1.1 Preselection of photons

Photons may begin to shower before reaching the ECAL and therefore not all of their energy is measured. The energy (and p_T) of each photon is corrected due to poor knowledge of the material effects upstream from the ECAL. The energy scale is restored by applying further η and ϕ dependent energy corrections that are determined from the well understood $Z \to ee$ resonances [47]. For converted photons, further corrections are made to the energy scale, from the radius of the conversion curvatures, which is not taken into account in the energy rescaling just mentioned.

In each event it is required that there be at least two loose photons with $p_T > 25\,\text{GeV}$ and $|\eta| < 2.37$. Due to poor reconstruction of photons between the barrel and the end-caps of the calorimetry system, photons are rejected in the $1.37 < |\eta| < 1.52$ region. Photons which pass through known dead regions of the calorimeter are also excluded.

Out of the preselected photons, the photon with the highest transverse momentum $p_{T,\gamma 1}$ is referred to as the leading photon and the photon with the second highest transverse momentum $p_{T,\gamma 2}$ is referred to as the subleading photon. After pre-selecting a leading and subleading photon, the cuts are tightened. The event is rejected if:

- $p_{T,\gamma 1} < 40\,\text{GeV}$ or if $p_{T,\gamma 2} < 30\,\text{GeV}$;
- the leading or subleading photon does not satisfy the tight quality cuts;
- the leading or subleading photon is not isolated.

The p_T and pseudorapidity measurements are corrected using the chosen PV.

5.1.2 Reweighting and Corrections Applied to MC

Although the MC simulates the signal well, there are some discrepancies when comparing data and MC. Each event in the MC signal samples described in Chapter 3 is reweighted to correct for the differences between data and MC. Reweighting is applied for:

- Pileup effects;
- Inconsistencies in the position of the beam spot;
- Interference (in the gluon-gluon fusion samples) from the $gg \to \gamma\gamma$ amplitude.

Additional treatment is applied to the energy measurements and the shower shape variables of the photons, such that the simulated detector effects are consistent with that of data.

5.1.3 Categorisation of $\gamma\gamma$ events

It was demonstrated by the ATLAS collaboration that it is benificial to divide the selected $\gamma\gamma$ events into categories based on the properties of the two leading p_T photons. Each category has different signal-background ratios and signal resolutions. The categories are weighted accordingly, which improves the overall signal sensitivity. For the late 2012 analysis produced by the ATLAS collaboration [47], additional categories were included to increase the sensitivity to the VBF, WH and ZH processes. This is also useful to study individual processes such as VBF.

As described previously, tag jets are present in VBF signal events and leptons or jets are present in WH and ZH signal events. This requires the identification of jets, electrons and muons. The overlap removal is done using the same procedure as described in Reference [47]. The two leading selected photons take preference over all other objects, which are selected in the following order:

- Electrons are selected if they are not overlapped with any of the two leading photons ($\Delta R_{e,\gamma} < 0.4$);
- Jets are selected if they are not overlapped with any of the two leading photons ($\Delta R_{j,\gamma} < 0.4$) or with any of the selected electrons ($\Delta R_{j,e} < 0.2$);
- Muons are selected if they are not overlapped with any of the two leading photons ($\Delta R_{\mu,\gamma} < 0.4$) or with any of the selected jets ($\Delta R_{\mu,j} < 0.4$).

A category is dedicated to be enriched in WH or ZH signal events where the W or Z boson decays leptonically. If the event contains at least one electron or at least one muon, the event is placed in the so called 'lepton category'. In the special case where a muon and an electron are separated by $\Delta\phi < 0.005$ or $\Delta\eta < 0.005$ this event is not be placed in the lepton category. A second category is dedicated to be enriched in WH or ZH signal events but where the W or Z decays hadronically. Events are placed in a so called low mass dijet (LMDJ) category[1] if:

- $|p_{T,t\gamma\gamma}| > 60\,\text{GeV}$ and
- $|\Delta\eta_{jj}| < 3.5$ and
- $60 < M_{jj} < 110\,\text{GeV}$.

Another jet category has been optimised so that it is rich in H$\rightarrow\gamma\gamma$ signal events which are produced by the VBF mechanism and at the same time reduces other signal and Standard Model backgrounds. Events are placed in this category if:

- $\Delta\eta_{jj} > 2.6$ and
- $\Delta\phi_{jj,\gamma\gamma} > 2.8$ and
- $M_{jj} > 400\,\text{GeV}$.

Due to the high invariant mass characteristic of the two tag jets, this category is therefore referred to as the high mass di-jet category (HMDJ). The remaining diphoton events are placed into a category, which is divided into sub-categories based on the photons $p_{T,t}$, pseudorapidty and conversion status. [56]. Since the signal events in this category are expected to be dominantly gluon-gluon fusion this category is named the gluon-gluon fusion enriched category (GGFE). As less than 1 ttH events are expected at $13\,\text{fb}^{-1}$ no category is designated for these signal events. The flow diagram describing the event categorisation is shown in Figure 5.1.

5.2 Motivation for the re-optimisation of the HMDJ category

The first $13\,\text{fb}^{-1}$ of 2012 ATLAS data and the $H\rightarrow\gamma\gamma$ signal MC samples for all production processes generated with a Higgs mass, $m_H = 125\,\text{GeV}$ were put through the event selection and

[1] This category is named for the event characteristics where the invariant mass of the two leading jets is in a mass window around the W and Z boson masses.

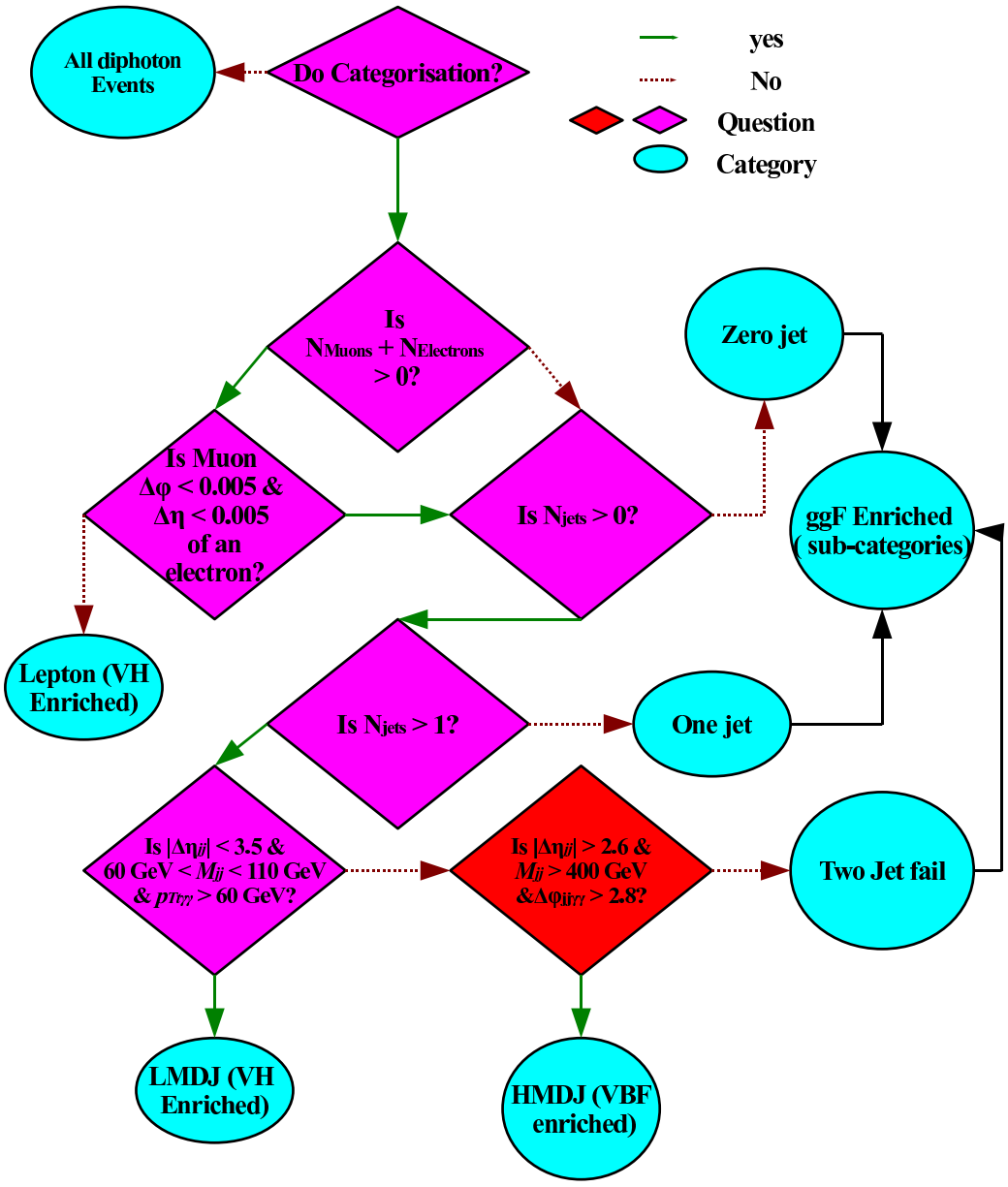


Figure 5.1: Flow chart of the nominal categorisation procedure described in the text.

categorisation as described in Section 5.1.3. The corrections described in Section 5.1.2 were applied to the MC and the data with the events in the signal region excluded. The total MC events have been scaled to obtain the signal yield at 13 fb^{-1} for each category. This is obtained by multiplying the expected number of events at 13 fb^{-1}, as shown in Table 3.1, by the selection efficiency. Since each event carries a weight to correct for pile-up and data-MC inconsistencies, the selection efficiency for each category and given signal process, p, is given by where n_p^c is the sum of all the event weights in category, c, for a given signal process and N_p is the summation of the weights of all events in the MC sample for the same process.

$$\varepsilon_p^c = \frac{n_p^c}{N_p} \tag{5.1}$$

where n_p^i is the summation of all the event weights in category c for a given signal process and N_p^c is the summation of total weighted events in the MC sample.

Using the categorisation procedure described in Section 5.1.3 the signal yields for 13 fb^{-1} of data are categorised, and shown in Table 5.1. Yields are shown for events with $100 < m_{\gamma\gamma} < 160$ GeV. For data, the signal region is excluded.

The statistical uncertainty on the data is a Poisson error and is therefore $\sqrt{n}$ where n is the number of data events selected for each category. The uncertainty on the yields is determined through the statistical uncertainty on the signal efficiency, which is obtained through error propagation [57]

$$\delta\varepsilon_p^c = \frac{\sqrt{\sum_+ w^2 (\sum_- w)^2 + \sum_- w^2 (\sum_+ w)^2}}{(\sum w)^2} \tag{5.2}$$

where

- $\sum w$ is the summation of all the event weights in the MC sample for a given process;
- $\sum_+ w^2$ is the summation of the square of all the MC event weights selected to a category c for a given process;
- $\sum_- w$ is the summation of all the MC event weights not selected to a category c for a given process;
- $\sum_- w^2$ is the summation of the square of all the MC event weights not selected to a category c for a given process;

Category	Data	ggF	VBF	WH	ZH	ttH
lepton	126	0.056	0.010	1.288	0.314	0.351
LMDJ	382	2.908	0.236	1.031	0.543	0.079
HMDJ	274	2.409	5.308	0.029	0.012	0.007
Two Jet fail	7476	24.15	5.041	2.311	1.394	0.735
One tag jet	16334	66.10	6.687	1.844	1.038	0.004
Zero tag jet	39159	132.7	1.530	0.724	0.765	0.004

Table 5.1: Weighted MC events in the range $100 < m_{\gamma\gamma} < 160\,\text{GeV}$, scaled to $13\,\text{fb}^{-1}$ for all Higgs production mechanisms. The scaling factors were calculated from the selection efficiency for each category and the cross sections and branching ratios shown in Table 3.1. The amount of selected data is also shown with events in the range $120 < m_{\gamma\gamma} < 130\,\text{GeV}$ removed, as these events will not be used to estimate the background.

Category	Data	ggF	VBF	WH	ZH	ttH
lepton	11.2	0.0043	0.0009	0.0388	0.0148	0.0086
LMDJ	19.5	0.0318	0.0045	0.0354	0.0193	0.0043
HMDJ	16.6	0.0287	0.0204	0.0058	0.0027	0.0012
Two jet fail	86.5	0.0901	0.0199	0.0511	0.0299	0.0118
One tag jet	127.8	0.1432	0.0224	0.0461	0.0259	0.0009
Zero tag jets	197.9	0.1892	0.0113	0.0295	0.0224	0.0010

Table 5.2: Statistical uncertainty on the event yields shown in Table 5.1.

- $\sum_{+} w$ is the summation of all the MC event weights selected to a category c for a given process;

The uncertainties are shown in Table 5.2

The lower section of Table 5.1 is the GGFE category which has been divided into new sub categories based on jet multiplicity:

- Selected events with two tag jets (as described in Section 3.1) that didnt fulfil LMDJ or HMDJ requirements (Two jet fail);
- Selected events with only one tag jet (One jet);
- Selected events with no tag jets (Zero jet).

5.31 VBF signal events are selected as HMDJ, nearly as many events have two tag jets but are otherwise failing the HMDJ requirements. In addition 2.41 gluon-gluon fusion signal events are also selected as HMDJ. This gluon-gluon fusion contamination is therefore reducing the purity of the VBF signal in this category. In anticipation of a cross section measurement of the gluon-gluon

fusion and VBF cross sections (presented in a later chapter) and given that VBF signal is already limited in a dataset of $13\,\text{fb}^{-1}$, it is desirable to increase the VBF signal selection efficiency and the purity of this category.

Optimisation metrics will now be defined, in which to quantify improvement in performance:

- VBF signal yield in the HMDJ category (N_{VBF}^{HMDJ}) at $13\,\text{fb}^{-1}$ in $100 < m_{\gamma\gamma} < 160\,\text{GeV}$;
- Gluon-gluon signal contamination in the HMDJ category (c_{ggF}^{HMDJ}) in $100\,\text{GeV} < m_{\gamma\gamma} < 160\,\text{GeV}$;
- Expected VBF signal significance (Z_{VBF}^{HMDJ}) for the HMDJ category.

To ensure the HMDJ is VBF enriched, gluon-gluon signal contamination in the HMDJ category should be minimised, this is defined as:

$$c_{ggF}^{HMDJ} = \frac{N_{ggF}^{HMDJ}}{N_{VBF}^{HMDJ} + N_{ggF}^{HMDJ}} \tag{5.3}$$

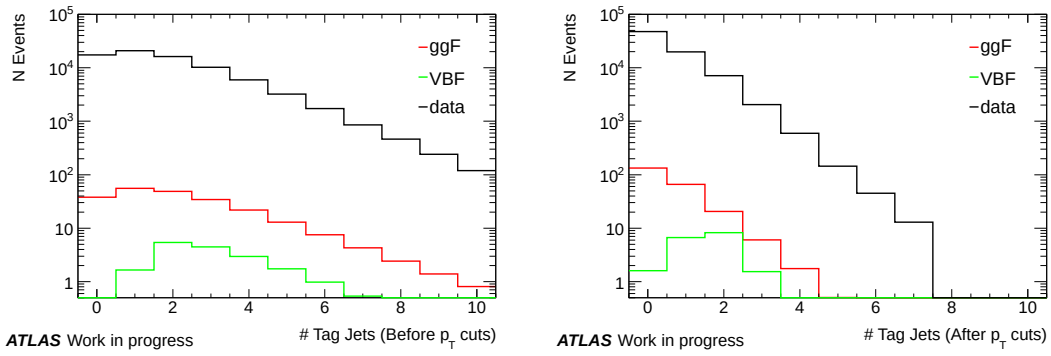
where N_{ggF}^{HMDJ} is the gluon-gluon fusion signal yield at $13\,\text{fb}^{-1}$ in $100 < m_{\gamma\gamma} < 160\,\text{GeV}$ which are categorised as HMDJ events. Expected VBF signal significance is also required to be maximised, so as to minimise the standard error on the signal. The significance metric is defined here as:

$$Z_{VBF}^{HMDJ} = \frac{N_{VBF}^{HMDJ}}{\sqrt{N_{VBF}^{HMDJ} + N_{ggF}^{HMDJ} + N_{data}^{SB,HMDJ}}} \tag{5.4}$$

where $N_{data}^{SB,HMDJ}$ is the number of data events in the sidebands (as defined in Section 3.3.2) and gluon-gluon fusion is treated as background.

5.3 Optimisation of the HMDJ Event Selection

As described in the preceding section, the current HMDJ event selection only captures a fraction of the VBF signal and contains a non-negligible contamination from gluon fusion events. The remainder of this chapter is devoted to investigating possible ways of improving the HMDJ selection.



(a) p_T thresholds are not applied

(b) p_T thresholds are applied

Figure 5.2: Number of tag jets identified in the background (data sidebands) and in the VBF and gluon-gluon fusion signal, (a) when the p_T thresholds are relaxed, and (b) when the p_T thresholds are applied

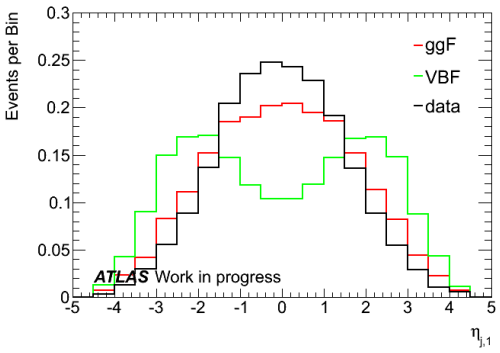
5.3.1 Optimisation by re-adjusting the tag jet p_T requirements

One of the reasons for the loss of VBF signal efficiency in the HMDJ category arises from one of the jets in a VBF event not being identified as a tag jet. The main cause of this is mostly due to tag jets failing the p_T thresholds of the tag jet definitions. This is demonstrated in Figure 5.2. The diphoton events are binned in accordance to the number of tag jets identified in the event (See Figure 5.2(b). The same is shown in Figure 5.2(a) except no p_T requirements have been applied in the tag jet definition. It is clear to see that in the VBF signal there are many events which have at least two tag jets, as is expected. When the minimum p_T requirements are applied to the tag jet definitions (25 GeV for $|\eta| < 2.5$ and 30 GeV for $|\eta| > 2.5$) the average number of tag jets decreases.

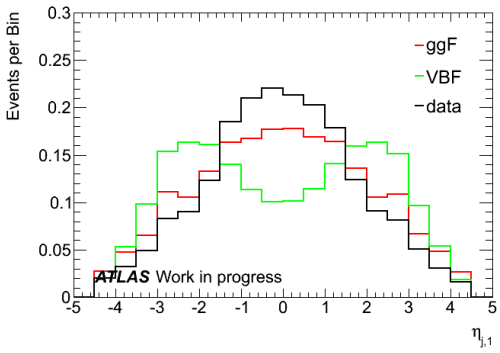
A possible way to increase the VBF signal selection efficiency in the HMDJ category, is to reduce the p_T thresholds on the tag jets. Three additional cut-base selections were proposed in addition to the nominal cut-based selection. For the first, the HMDJ category is defined in the nominal way, except the tag jets are instead required to have a p_T >25 GeV in both the barrel ($|\eta| < 2.5$) and in the endscaps ($|\eta| > 2.5$). The second is the same as the first, now the tag jet p_T threshold is reduced to p_T >20 GeV and the third is P_T >15 GeV. N_{VBF}^{HMDJ}, Z_{VBF}^{HMDJ} and c_{ggF}^{HMDJ} were calculated for each of these redefined HMDJ category defined by the 3 additional cut based proposals. By lowering the tag p_T thresholds the expected VBF signal can increase (see Table 5.3). This only does however achieve a moderate increase in significance if at all, and always results in increased contamination from gluon-gluon fusion signal.

p_T Thresholds	N_{VBF}^{HMDJ}	Z_{VBF}^{HMDJ}	c_{ggF}^{HMDJ}
$p_T > 25(30)$ GeV for $\|\eta\| < 2.5(> 2.5)$	5.308	0.316	0.312
$p_T > 25$ GeV	5.874	0.324	0.334
$p_T > 20$ GeV	6.522	0.312	0.377
$p_T > 15$ GeV	7.014	0.285	0.430

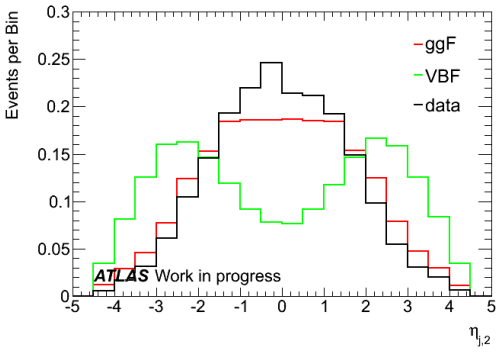
Table 5.3: Expected VBF signal, VBF significance and gluon-gluon fusion contamination in the HMDJ category, for different definitions (in terms of p_T thresholds) of the tag jets. The nominal p_T cuts are compared with alternative scenarios with lower p_T cuts, as described in this section.



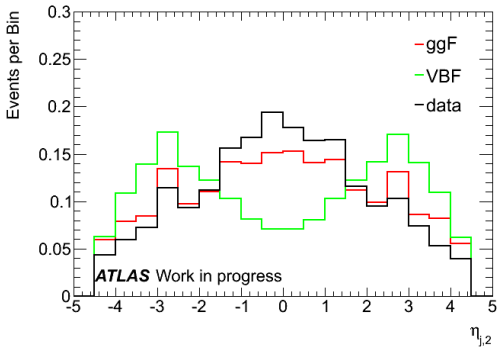
(a) η_{j1} distribution for nominal tag jets



(b) η_{j1} distribution for tag jets selected with $p_T > 15$ GeV



(c) η_{j2} distribution for nominal tag jets



(d) η_{j2} distribution for tag jets selected with $p_T > 15$ GeV

Figure 5.3: η distributions of the tag jets, for the background (data sidebands) and the VBF and gluon-gluon fusion signals. η distributions of the highest-p_T selected tag jet, using (a) the nominal p_T thresholds and (b) the lower p_T threshold of 15 GeV. η distributions of the second highest-p_T selected tag jet, using (c) the nominal p_T thresholds and (d) the lower p_T threshold of 15 GeV.

Lowering the tag jet p_T thresholds results in increased background acceptance as shown by the reduction of Z_{VBF}^{HMDJ}. The likelihood of selecting a forward low p_T jet from pileup will also increase. This is demonstrated in Figure 5.3, where the multiplicity of identified tag jet candidates as a function of η is shown for nominal p_T thresholds and for the lower threshold of 15 GeV. In addition, lowering the p_T thresholds opens up a region of phase space with large systematic uncertainties such as the uncertainty on the jet energy scale calibration (JES).

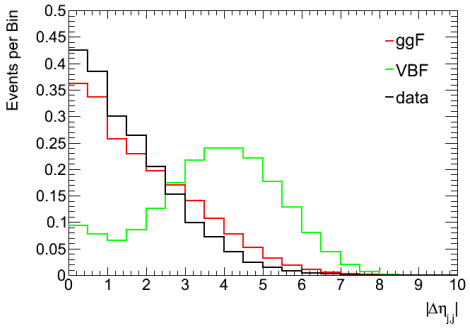
5.4 Optimisation using a multi-variate classifier

An alternative option for the reoptimisation of the HMDJ is to recover events that have migrated into the 2 jet fail category by replacing the red diamond box in Figure 5.1 with a multi-variate analysis (MVA) classifier. In particular a boosted decision tree has been investigated. An MVA classifier requires information from many variables and will make a decision based on the information of all the variables combined. Before proceeding to using a BDT classifier, input variables to the BDT will be decided upon.

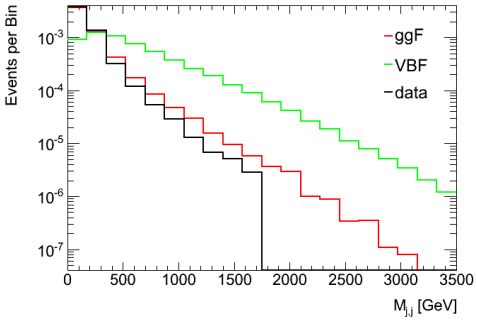
5.4.1 Choice of Input Variables for the HMDJ BDT Classifier

The purpose of reoptimisation in this analysis is to improve the VBF signal efficiency and simultaneously reduce the background selected and the contamination from the gluon-gluon fusion signal. The signal-background separation has been investigated for several variables. Most of the variables relate to the properties of the tag jets, as the tag jets are one of the main features of the VBF signal. Prior to selecting which variable to use as input to the BDT, the variables were grouped into different types. 'Type A' are those for which there is a distinctive separation between VBF signal on one hand and the background and gluon-gluon fusion on the other. These variables are: $\Delta\eta_{jj}$; M_{jj}; η_{j1}; η_{j2}; $\eta_{j1}.\eta_{j2}$; $p_{T,j1}$; $p_{T,j2}$ and $|\vec{p_{T,\gamma\gamma}} + \vec{p_{T,jj}}|$ (see Figure 5.4).

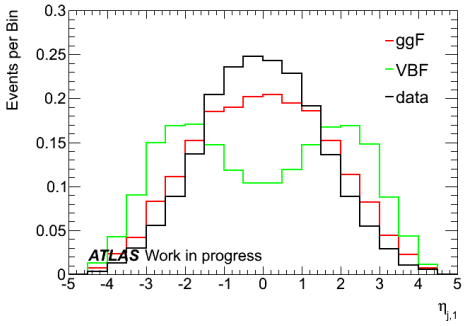
'Type B' variables are those which have good separation of VBF signal from background but the variable distributions of gluon-gluon fusion signal is more similar to that of the VBF signal, these are: $\Delta\phi_{\gamma\gamma,jj}$; $p_{T,\gamma 1}$; $\Delta R_{\gamma 1,j1}$ and $p_{T,t\gamma\gamma}$ (see Figure 5.5). The variables that were initially chosen for the BDT classifier were all type A variables, M_{jj}, η_{j1} (the absolute pseudorapidity of the leading jet), η_{j2} (the absolute pseudorapidity of the subleading jet), $p_{T,j1}$ (the transverse momentum of the leading jet), $p_{T,j2}$ (the transverse momentum of the subleading jet) and the p_T



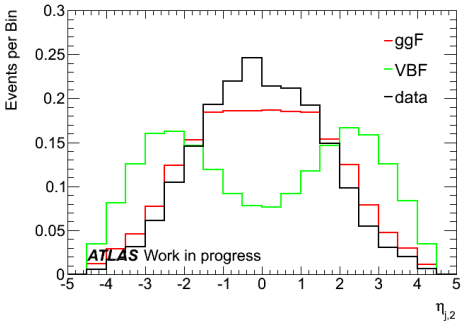
(a) $\Delta\eta_{jj}$



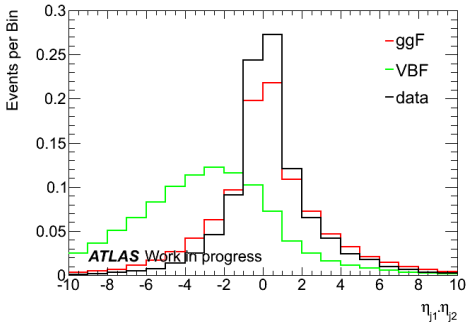
(b) M_{jj}



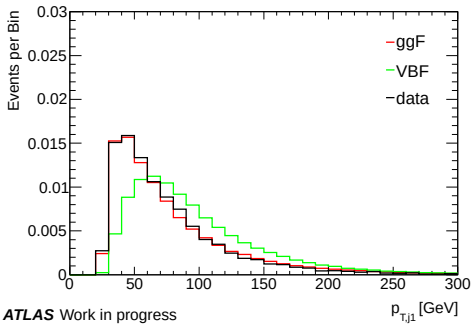
(c) η_{j1}



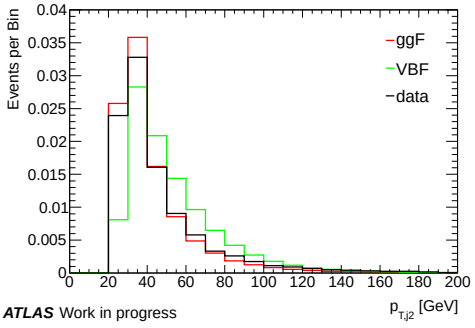
(d) η_{j2}



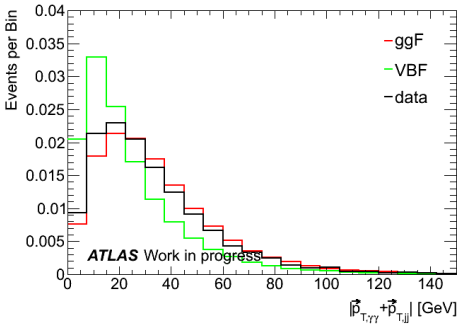
(e) $\eta_{j1}.\eta_{j2}$



(f) $p_{T,j1}$

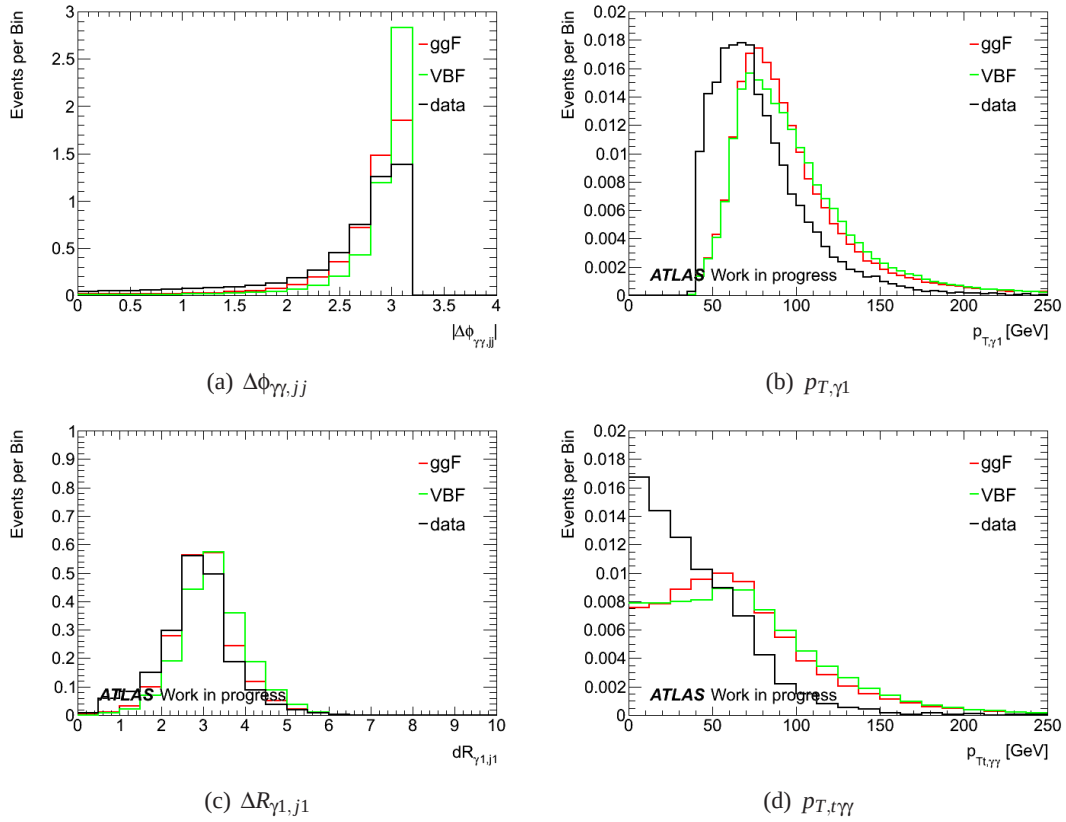


(g) $p_{T,j2}$



(h) $|\vec{pT,\gamma\gamma} + \vec{pT,jj}|$

Figure 5.4: Distributions of the 'Type A' variables: Those which offer a good discrimination between VBF signal on one side, and the background and the gluon-gluon fusion signal on the other. Events shown are those which have two photon candidates and two tag jet candidates, which are not categories as LMDJ.



(a) $\Delta\phi_{\gamma\gamma,jj}$

(b) $p_{T,\gamma 1}$

(c) $\Delta R_{\gamma 1,j1}$

(d) $p_{T,t\gamma\gamma}$

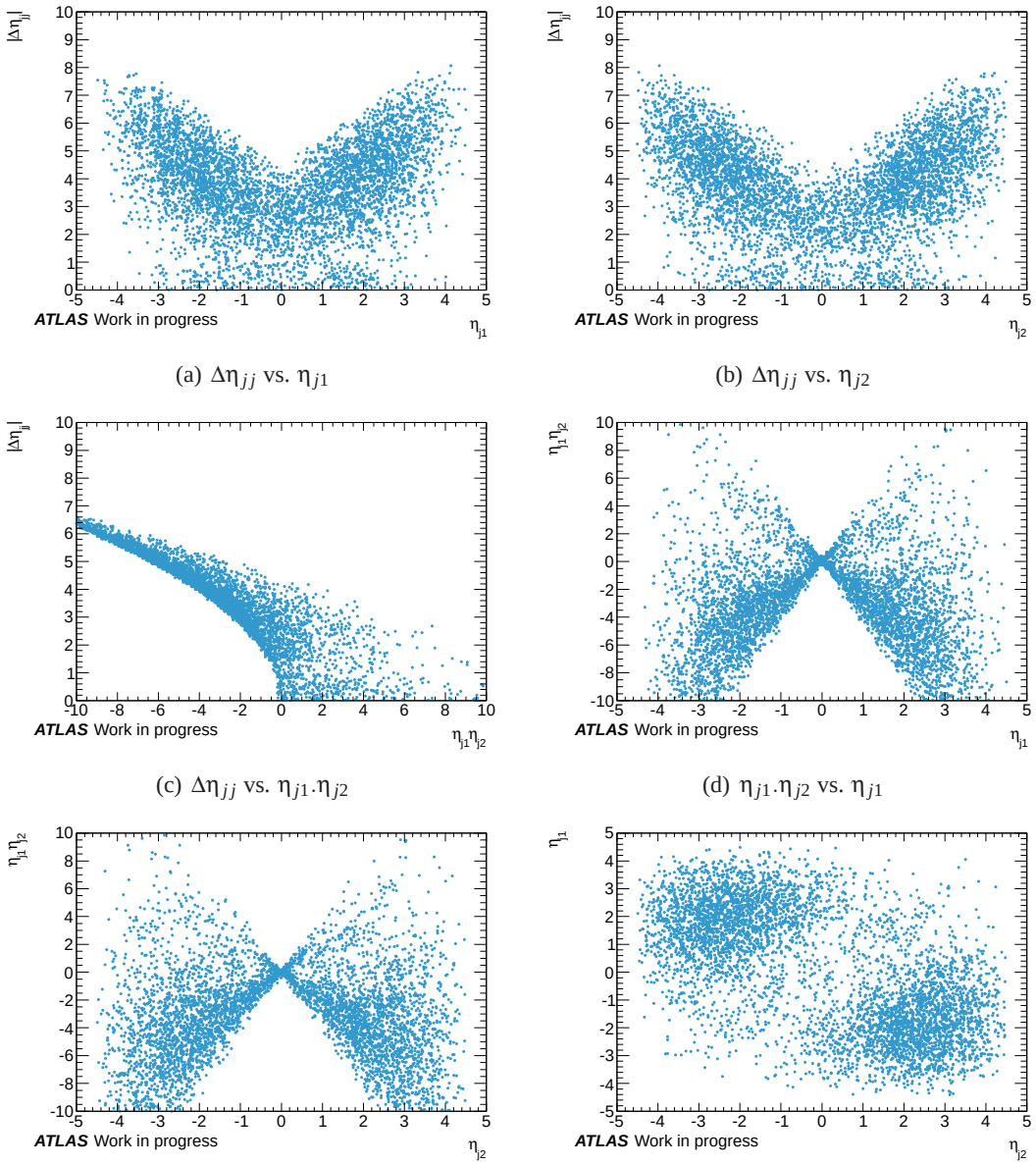
Figure 5.5: Distributions of the ‘Type B’ variables. Those which offer a good discrimination between VBF signal and the background but distributions of gluon-gluon fusion signal is more similar to that of the VBF signal. Events shown are those which have two photon candidates and two tag jet candidates, which are not categories as LMDJ.

balance variable $|\vec{p_{T\gamma\gamma}} + \vec{p_{T,jj}}|$.

Although $\eta_{j1}.\eta_{j2}$ and $\Delta\eta_{jj}$ appear to be strong discriminating variables, these variables are highly correlated with η_{j1} and η_{j2} as shown in Figure 5.6. It is predicted that $\eta_{j1}.\eta_{j2}$ and $\Delta\eta_{jj}$ in addition with η_{j1} and η_{j2} variables would add no extra discriminating power to the MVA, so just η_{j1} and η_{j2} were chosen for the initial baseline training. In summary the following Type A variables will be considered: η_{j1}, η_{j2}, M_{jj}, $p_{T,j1}$, $p_{T,j2}$ and $|\vec{p_{T\gamma\gamma}} + \vec{p_{T,jj}}|$. Type B variables will be added or removed later on, to see if any extra separation power can be gained; this will be shown in the later sections.

By using the $\Delta\phi_{\gamma\gamma,jj}$ variable there is a potential for a systematic error. It was discovered by the ATLAS collaboration that there is an uncertainty in the MC modelling of the difference in azimuthal angle between the two tag jets. The uncertainty arises in the analysis for $\Delta\phi_{\gamma\gamma,jj} > 2.94$. To remove any potential bias $\Delta\phi_{\gamma\gamma,jj}$ is set to 2.95 for $\Delta\phi_{\gamma\gamma,jj} > 2.94$

It is known that BDTs have the advantage that adding weak or correlated variables to the classifier does not degrade the performance of the classifier [58]; this will be demonstrated to be the case in Section 5.4.5. Nevertheless having a large number of input variables in the BDT increases the chance of there being large associated systematic uncertainties.



(a) $\Delta\eta_{jj}$ vs. η_{j1}

(b) $\Delta\eta_{jj}$ vs. η_{j2}

(c) $\Delta\eta_{jj}$ vs. $\eta_{j1}.\eta_{j2}$

(d) $\eta_{j1}.\eta_{j2}$ vs. η_{j1}

(e) $\eta_{j1}.\eta_{j2}$ vs. η_{j2}

(f) η_{j1} vs. η_{j2}

Figure 5.6: Scatter plots showing correlations between η_{j1}, η_{j2}, $\eta_{j1}.\eta_{j2}$ and $\Delta\eta_{jj}$ for events with two photon candidates and two tag jet candidates, which are not categorised as LMDJ.

	M_{jj} [GeV]	$\Delta\phi_{\gamma\gamma jj}$	$\Delta\eta_{jj}$
Event 1	400	2.6	3.4
Event 2	401	2.4	3.6

Table 5.4: The properties of two Hypothesised events that will go through an example BDT. The value of M_{jj}, $\Delta\phi_{\gamma\gamma jj}$ and $\Delta\eta_{jj}$ is shown for each event.

5.4.2 Boosted Decision Trees (BDT)

A multivariate analysis (MVA) technique has been used to improve the signal to background separation with respect to the nominal cut-based analysis. In particular, the use of decision trees has been investigated.

The schematic in Figure 5.7 will be used to demonstrate the selection of two example signal events, where the quantity of each variable is shown in Table 5.4. The goal is to classify any given event as a signal or a background candidate.

The event begins at the root node. In the example schematic, the events can follow one of two branches depending on whether $M_{jj} > 300$ GeV or not. Using Event 1 as an example, the condition at the root node is satisfied and the event is accepted via branch B, on the right. At the end of each branch, there is a node where another cut is applied. This is repeated until a node stops branching, at which point the node is referred to as a leaf. If an event lands on a signal leaf it is classified as signal and if it lands on a background leaf it is classified as background.

The advantage of the decision tree approach over the standard cut-based selection, is demonstrated using the example signal events in Table 5.4. Both of these events would be classified as signal by the decision tree, however, Events 1 and 2 have different outcomes from the $\Delta\phi_{\gamma\gamma jj}$ cut on the node at the end of branch B. In the standard cut based approach, this would have resulted in Event 2 being rejected. However event 2 is recovered in this example via branch H.

Training the Decision Tree

To construct a decision tree, training samples of signal and background events are required. 100 background events and 100 signal events were used in the example in Figure 5.7.

A signal or background leaf is classified as such based on the signal purity. The purity is calculated at the end of each branch,

$$p = \frac{n_s}{n_s + n_b} \tag{5.5}$$

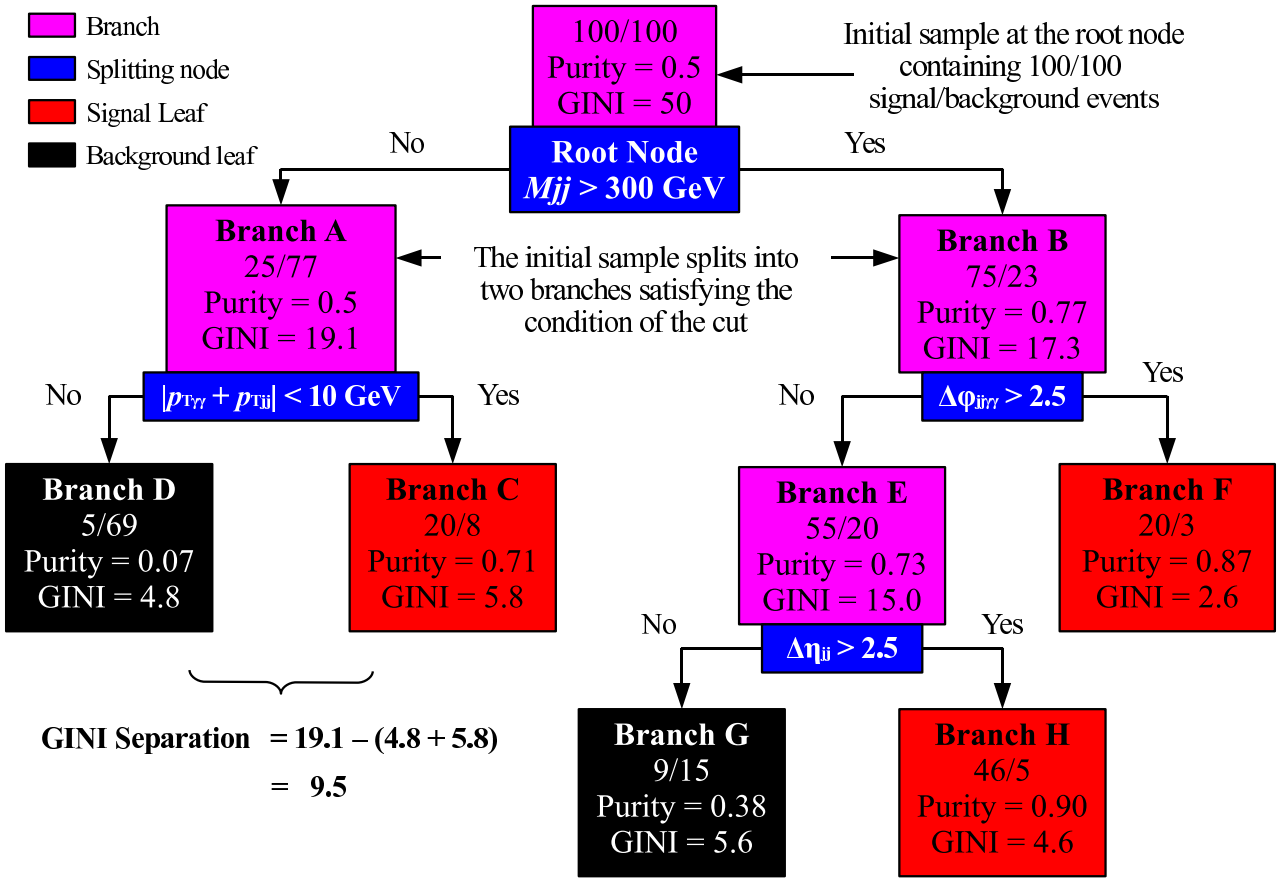


Figure 5.7: Schematic of a boosted decision tree, used to classify an events as signal or background.

where n_s is the number of training signal events accepted via a given branch and n_b is the number of background events accepted via a given branch.

To quantify the signal-background separation of either branch, the so called GINI index

$$\text{GINI} = (n_s + n_b)(p(1-p)) \tag{5.6}$$

is calculated for the daughter nodes at the end of each branch and parent node. The best signal-background separation is achieved when the difference in the GINI index of the parent node and the sum of indices of the two daughter nodes is maximised [59]. Based on this, in the training process, a variable is chosen which to cut on at a given node. The algorithm tests the GINI separation by applying a series of consecutively tighter cuts on the given selection variable. The number of different cuts applied is specified by the user. The chosen cut position is the one that gives the best GINI separation. Two nodes are formed from this cut and the above repeated until the number of signal events or background events on each branch falls below a threshold. At that point the branch is labelled signal or background, respectively.

Testing and Overtraining

If a classifier becomes far too complex, it becomes vulnerable to statistical fluctuations. The classifier can effectively learn individual signal and background events in a given training sample. If the classifier is applied to an independent testing sample, the same classification performance is not achieved. In fact it is more likely that more background will be classified as signal; this is referred to as overtraining. Decision trees are particularly vulnerable to overtraining due to the large number of nodes and the complexity of the tree structure, if for instance, there are too few data points relative to the number of nodes.

Boosting

Boosting is a way of stabilising the classifier by producing many trees and combining them together. Trees are produced iteratively where the event in the training sample carries an event weight, w. The purity in Equation 5.6 is modified to

$$p = \frac{\sum_{s=1}^{n_s} w_s}{\sum_{s=1}^{n_s} w_s + \sum_{b=1}^{n_b} w_b} \tag{5.7}$$

where w_s represents the weights of the signal events and w_b represents the weights of the background events. Likewise the GINI index is modified to

$$\text{GINI} = (\sum_{s=1}^{n_s} w_s + \sum_{b=1}^{n_b} w_b)(p(1-p)) \tag{5.8}$$

where N is the total number of signal and background events in the training sample. After the tree is constructed, the weight of each event is modified (boosted) depending on the event classification. If an event was correctly classified the event weight is unchanged and if the event was incorrectly classified the event weight is increased, thus defining a new training sample; another tree is then constructed. The motivation behind this is that the new tree will be more sensitive to the misclassified events.

Certain conditions will now be defined to quantify the amount of boosting an event receives:

- y_i: For true signal events $y_i = 1$ and for true background events $y_i = -1$;
- T_i^m: The classification of the i^{th} event by the m^{th} tree. Classification as signal is $T_i^m = 1$ and classification as background is $T_i^m = -1$;
- $I(y_i \neq T_i^m)$: Is a Boolean condition where $I(y_i \neq T_i^m = 1)$ if $y_i = T_i^m$ and 0 otherwise.

For the m^{th} tree the weight of each event is modified by

$$w_i \rightarrow w_i e^{-\xi I(y_i \neq T_i^m)} \tag{5.9}$$

where ξ is the learning rate of the BDT. Note that if $y_i = T_i^m$ then w_i is left unchanged. The weights are renormalised and form a new sample for the $(m+1)^{th}$ tree.

The type of boosting used in this analysis is the gradient boost. In this case ξ is a constant of $O(0.01)$ [59]. The smaller the value of ξ, the more robust the BDT is against outlier events. However, by making ξ small, many more iterations (number of trees) are required [58]. After many iterations each event is given a score based on the outcome of each tree. The score for each event is

$$T_i = \sum_{m=0}^{N_{Trees}} \xi T_i^m. \tag{5.10}$$

This score is the quantity which is cut on, which is referred to as the BDT response in Figure 5.8.

5.4.3 HMDJ BDT training procedure with TMVA

A BDT classifier is constructed using the TMVA software package [58]. Sideband data are used to model the background. Events in the signal region are not used to prevent any bias from genuine signal events. The VBF and gluon-gluon fusion signal events were generated at a Higgs boson mass $m_H = 125\,\text{GeV}$. The signal and background samples were both split equally into statistically independent training and testing samples, using a random splitting procedure provided by the TMVA software. The signal and background training samples are used to train the classifier and the testing sample is used to verify that there is no overtraining and to independently establish the actual performance of the BDT. A BDT classifier is constructed using the input variables outlined in Section 5.4.1. After the BDT has been trained various control plots are produced to monitor its performance. As an example, the output of a BDT with 6 variables M_{jj}, η_{j1}, η_{j2}, $p_{T,j1}$, $p_{T,j2}$ and $|\vec{p}_{T,\gamma\gamma} + \vec{p}_{T,jj}|$ is shown in Figure 5.8. As signal events generally receive a higher score, T_i, (also referred to as the BDT response) the testing and training signal events appear on the right of this plot and the background generally receives a lower score. The testing and training background events appear on the left of this plot. These distributions suggest that the classifier is not overtrained as the testing and training distributions are compatible within their uncertainties. This can be quantified using the Kolmogorov-Smirnov (KS) statistic, which is the probability that the two distributions are consistent with the same true parent distribution. An analysis performed by the ATLAS collaboration recomended that the KS should be greater than 0.1, in order to ensure no overtraining has occurred.

The classifier defines signal and background by placing a cut on T_i, which is determined by the user. An event with T_i greater than the cut (T_{cut}) is classified as signal, and as background otherwise. The choice of the cut will determine how much VBF signal, background and also the amount of gluon-gluon fusion signal that is selected into the HMDJ category. The choice of this cut can be investigated using the plots shown in Figure 5.9. Each point along the curves represents a different working point as the T_{cut} slides across the spectrum of events T_i in steps of 0.01 between -1 and 1. The blue band around the curve represents the statistical uncertainty on each metric. The uncertainty on the number of signal and background events is explained in Section 5.2. The uncertainty on the significance and the gluon-gluon fusion contamination were determined through error propagation.

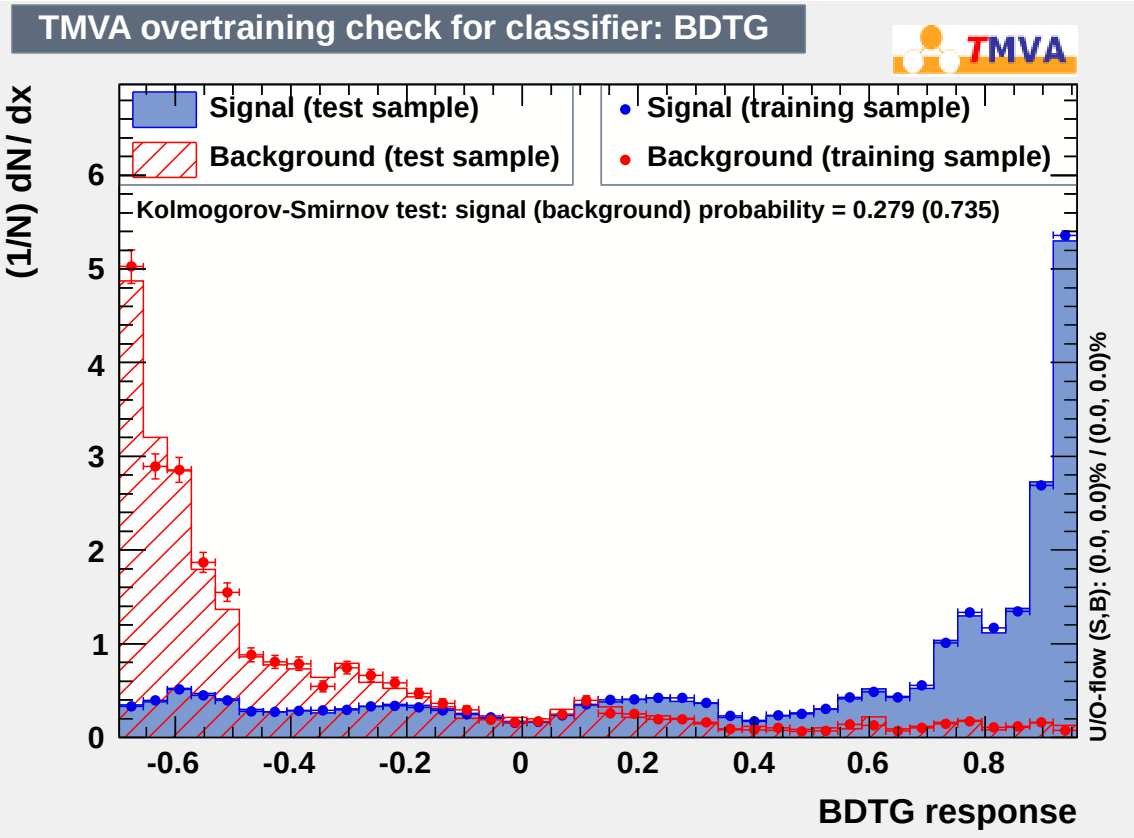


Figure 5.8: BDT response distributions of each event (T_i) for a BDT based on 6 discriminant variables M_{jj}, η_{j1}, η_{j2}, $p_{T,j1}$, $p_{T,j2}$ and $|\vec{p}_{T,\gamma\gamma} + \vec{p}_{T,jj}|$. The distributions are shown for the signal training and testing samples (blue dots and blue solid histogram, respectively) and for the background training and testing samples (red dots and red hashed histogram, respectively).

The curve of N_{VBF}^{HMDJ} against Z_{VBF}^{HMDJ} from the training samples is shown in Figure 5.9(a), N_{VBF}^{HMDJ} against c_{ggF}^{HMDJ} from the training sample is shown in Figure 5.9(b). The response of the classifier to the independent testing sample for N_{VBF}^{HMDJ} against Z_{VBF}^{HMDJ} is shown in Figure 5.9(c) and N_{VBF}^{HMDJ} against $c_{ggF}^{HMDJ}$is shown in Figure 5.9(d). Both the testing and training samples have suggested that a clear improvement can be gained with respect to the nominal cut-based analysis. Either way, improvement is gained with respect to the nominal cut-based analysis. If a working point is chosen that yields the same signal selection efficiency (N_{VBF}^{HMDJ}) as the nominal cut-based analysis 8.2% improvement on VBF signal significance Z_{VBF}^{HMDJ} is gained. Or, if a working point was chosen that yielded the same VBF signal significance Z_{VBF}^{HMDJ} as the nominal cut-based analysis one would achieve a 21.5% improvement on the VBF signal yield. Lowering the p_T thresholds which define the tag jets, does have a comparable improvement with respect to VBF signal yield, however the gluon-gluon fusion signal contamination in the HMDJ category is much higher than the BDT classifier would yield, as shown in Figures 5.9(b) and 5.9(d).

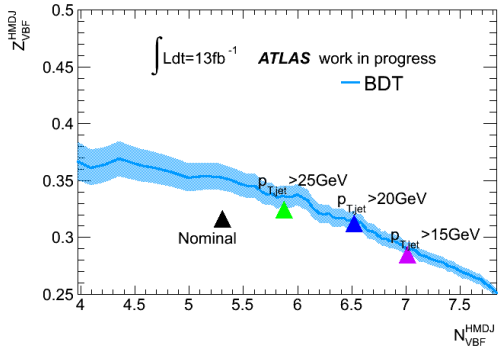
5.4.4 Internal parameters of the BDT

There are several internal parameters of the BDT that can potentially be adjusted, to enhance the classification performance. The values that were recommended by ATLAS will be used throughout the rest of this thesis but some parameters will be investigated to check that the choice of value for each parameter will not cause any instability to the BDT performance. The parameters investigated were:

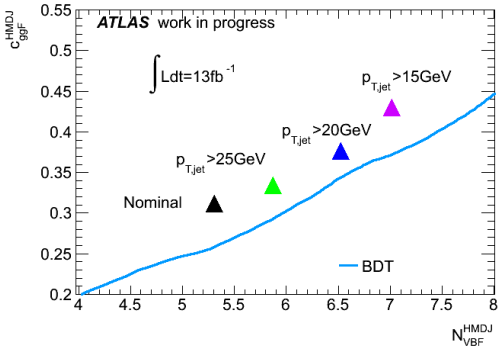
- Learning rate of the BDT ξ, also referred to as the shrinkage.
- Number of trees (NTrees);
- Minimum event number threshold on a branch (NEventsMin);
- The number of cuts tested to maximise the GINI separation between branches (NCuts);

The values recommended by ATLAS are shown in bold in Table 5.5. Each time a parameter was adjusted, all other parameter were fixed to the values shown in bold. The BDT was re-trained and the performance in terms of N_{VBF}^{HMDJ} and Z_{VBF}^{HMDJ} was determined for a variety of T_{cut} values.

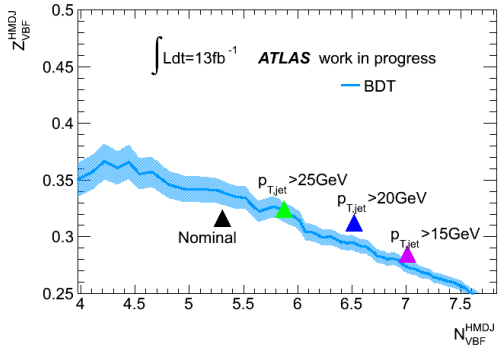
NEventsMin can potentially cause overtraining if it is set too small. Allowing too few events on a signal or background leaf would increase the number of nodes and the tree becomes too



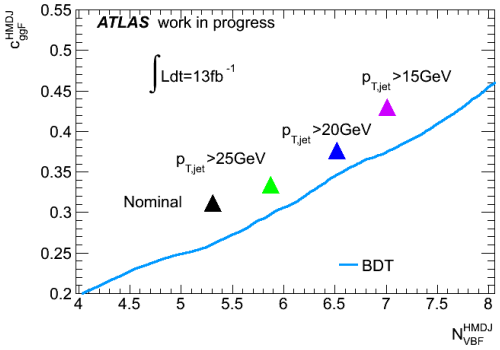
(a) Z_{VBF}^{HMDJ} vs N_{VBF}^{HMDJ} training sample



(b) c_{ggF}^{HMDJ} vs N_{VBF}^{HMDJ} training sample



(c) Z_{VBF}^{HMDJ} vs N_{VBF}^{HMDJ} testing sample

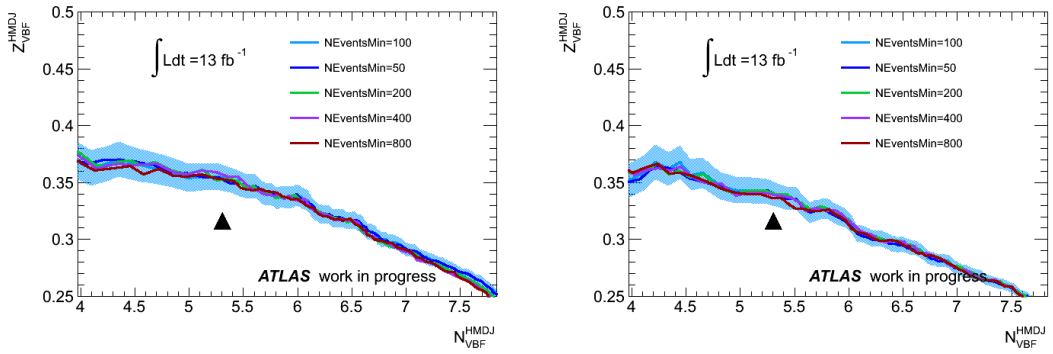


(d) c_{ggF}^{HMDJ} vs N_{VBF}^{HMDJ} testing sample

Figure 5.9: Performance of a BDT classifier, trained with variables M_{jj}, η_{j1}, η_{j2}, $p_{T,j1}$, $p_{T,j2}$ and $|\vec{p}_{T,\gamma\gamma}+\vec{p}_{T,jj}|$ compared with the performance of the nominal cut-based selection of the HMDJ and potential changes to the cut-based selection, involving lowering the p_T thresholds of the tag jets.

Parameter	Internal parameter values					
Shrinkage	0.025	**0.05**	0.1	0.2	0.4	0.8
NTrees	200	600	**1000**	1400		
NEventsMin	50	**100**	200	400	800	
NCuts	10	**30**	50	70	90	

Table 5.5: Study of different values used for the internal configuration of the BDT (recomended values are shown in bold). Each time a parameter is adjusted, the other parameters are fixed to the recomended values.



(a) Training Sample (b) Testing Sample

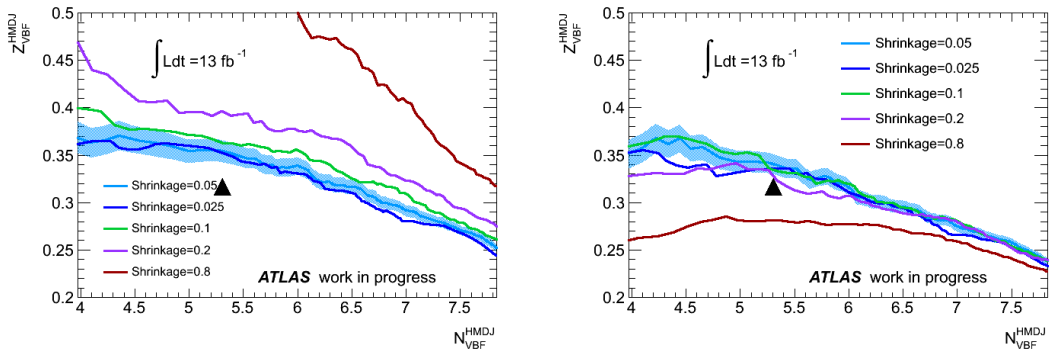
Figure 5.10: Performance in terms of N_{VBF}^{HMDJ} and Z_{VBF}^{HMDJ} investigated for NEventsMin ranging between 50 and 800 events and compared with the nominal performance indicated by the black triangle. Each value was tested on a BDT based on 6 discriminant variables M_{jj}, η_{j1}, η_{j2}, $p_{T,j1}$, $p_{T,j2}$ and $|\vec{p}_{T,\gamma\gamma}+\vec{p}_{T,jj}|$. All other internal parameters were set to the values recommended by ATLAS.

complex. On the other hand, if NEventsMin is too large the BDT would be too simple and the performance would degrade. NEventsMin was investigated using the values shown in Table 5.5. The performance from the training is shown in Figure 5.10(a) and on the independent testing sample in Figure 5.10(b). There appears to be no gain, loss or instability in the choice of either of these values.

The robustness of the BDT is predicted to be best providing ξ is kept at a small value. If ξ is too large the boosting will become too sensitive to the misclassified events and overtraining can occur. The performance is shown in Figure 5.11. The performance of the training is seen in Figure 5.11(a), which appears to increase with ξ. However when this is tested on an independent training sample the opposite effect occurs, which can be seen in Figure 5.11(b).

This is a clear example of overtraining. This is shown in the KS statistic, which is extremely low relative to the other values of ξ (see Table 5.6). By eye, it is easy to see that the testing and training distributions of T_i differ significantly for background, which is shown in Figure 5.12, especially at high T_i and low T_i.

The NCuts parameter was investigated to see whether having a finer granularity improves the GINI separation at each branch. Using the same BDT as before and setting the other parameters to the recommended values, N_{VBF}^{HMDJ} and Z_{VBF}^{HMDJ} were determined for a variety of T_{cut} and the performance of training and testing are shown in Figure 5.13. It is safe to assume that changing NCuts adds no extra performance.



(a) Training Sample (b) Testing Sample

Figure 5.11: Performance in terms of N_{VBF}^{HMDJ} and Z_{VBF}^{HMDJ} investigated for ξ ranging between 0.025 and 0.4. Each value was tested on a BDT based on 6 discriminant variables M_{jj}, η_{j1}, η_{j2}, $p_{T,j1}$, $p_{T,j2}$ and $|\vec{p}_{T,\gamma\gamma}+\vec{p}_{T,jj}|$. All other internal parameters were set to the values recommended by ATLAS.

Parameter	Value	KS	
		signal	background
Shrinkage	0.025	0.184	0.820
	0.05	0.279	0.7735
	0.1	0.0881	0.820
	0.2	0.0592	0.0194
	0.4	0.0507	0.000148
	0.8	5.06×10^{-3}	4.78×10^{-10}
NTrees	200	0.388	0.992
	600	0.344	0.904
	1000	0.279	0.735
	1400	0.219	0.332
	1800	0.155	0.780
NEventsMin	50	0.298	0.760
	100	0.279	0.735
	200	0.354	0.956
	400	0.439	0.982
	800	0.324	0.869
NCuts	10	0.0662	0.0741
	30	0.134	0.159
	50	0.0509	0.258
	70	0.0671	0.497
	90	0.158	0.704

Table 5.6: KS statistics for a BDT trained and tested with variables using variables M_{jj}, η_{j1}, η_{j2}, $p_{T,j1}$, $p_{T,j2}$ and $|\vec{p}_{T,\gamma\gamma}+\vec{p}_{T,jj}|$. The KS shown are for different internal configurations. The values in bold are the recommended values, each time a parameter in internal configuration is changed the other are fixed to the recommended value (shown in bold).

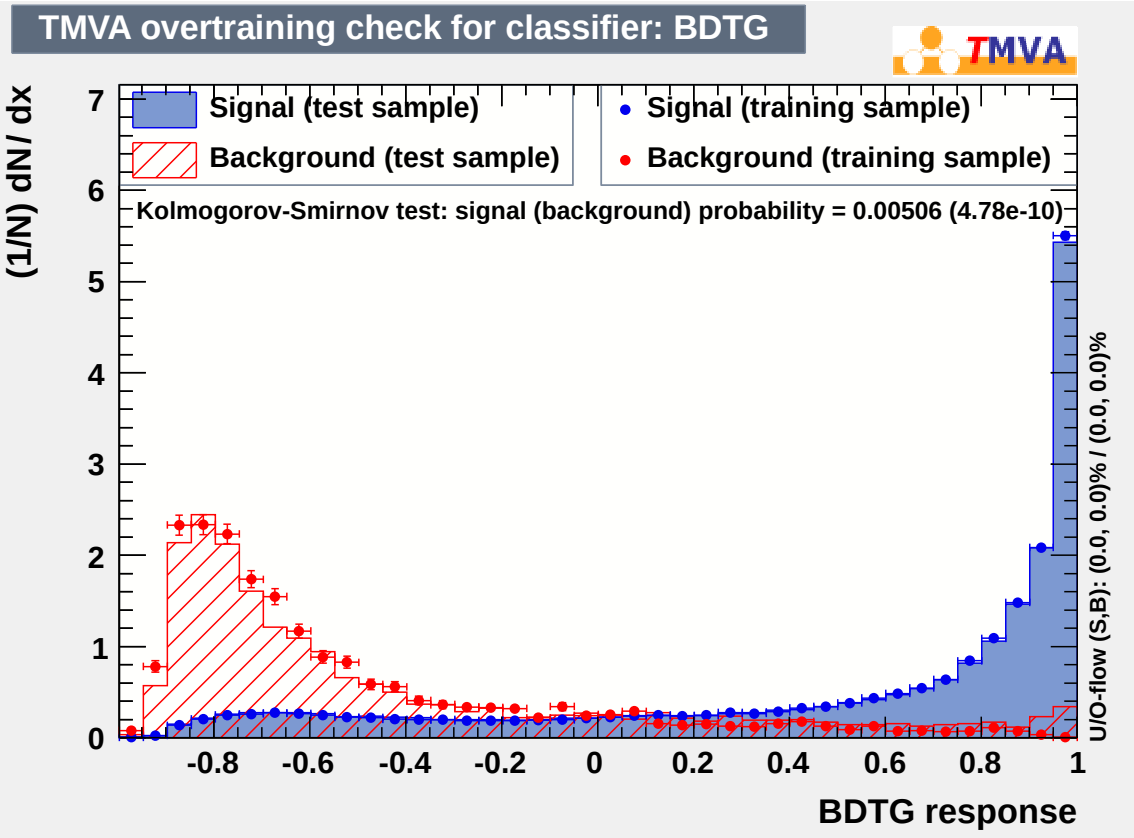
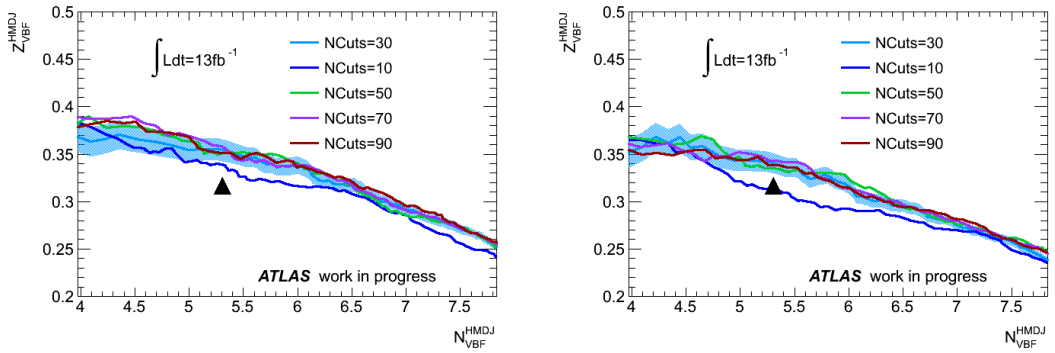


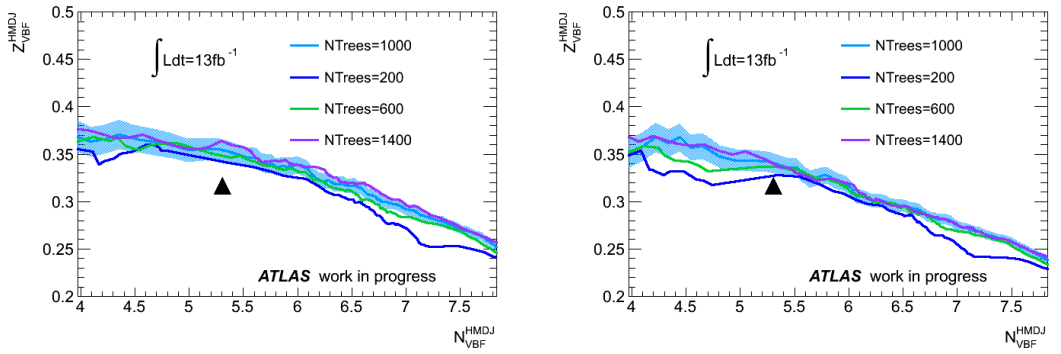
Figure 5.12: BDT response for each event (T_i) using variables M_{jj}, η_{j1}, η_{j2}, $p_{T,j1}$, $p_{T,j2}$ and $|\vec{p}_{T,\gamma\gamma}+\vec{p}_{T,jj}|$. The parameters in the internal configuration are set to the recommended values except ξ, which is set to 0.8.



(a) Z_{VBF}^{HMDJ} vs N_{VBF}^{HMDJ} training sample

(b) Z_{VBF}^{HMDJ} vs N_{VBF}^{HMDJ} testing sample

Figure 5.13: Performance in terms of N_{VBF}^{HMDJ} and Z_{VBF}^{HMDJ} investigated for NCuts ranging between 10 and 90. Each value was tested on a BDT training using variables M_{jj}, η_{j1}, η_{j2}, $p_{T,j1}$, $p_{T,j2}$ and $|\vec{p}_{T,\gamma\gamma}+\vec{p}_{T,jj}|$. All other internal parameters were set to the values recommended by ATLAS.



(a) Z_{VBF}^{HMDJ} vs N_{VBF}^{HMDJ} testing sample

(b) Z_{VBF}^{HMDJ} vs N_{VBF}^{HMDJ} testing sample

Figure 5.14: Performance in terms of N_{VBF}^{HMDJ} and Z_{VBF}^{HMDJ} investigated for NTrees ranging between 200 and 1800. Each value was tested on a BDT training using variables M_{jj}, η_{j1}, η_{j2}, $p_{T,j1}$, $p_{T,j2}$ and $|\vec{p}_{T,\gamma\gamma}+\vec{p}_{T,jj}|$. All other internal parameters were set to the values recommended by ATLAS.

As the recommended choice of values for each parameter is shown to be stable and optimal, these values will be adopted throughout the rest of this thesis.

5.4.5 Effects of Weak and η Variables on the performance of the BDT

In this section investigations are presented to

- re-visit section 5.4.1 and determine the best choice of tag jet η variables to use in the BDT;
- show that adding weak variables will not affect the performance of the BDT.

Choice of η variables

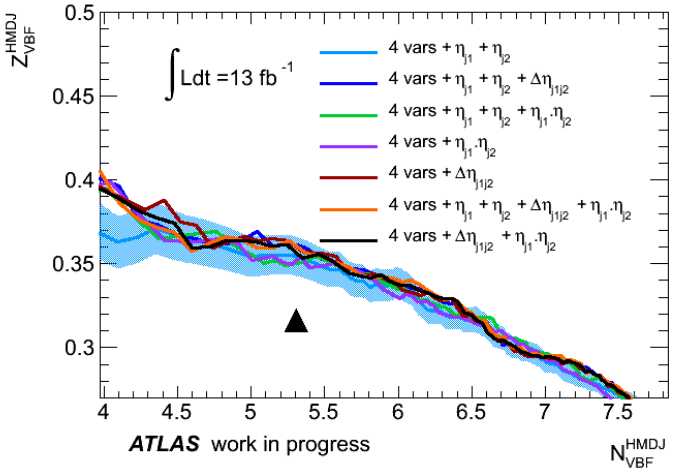
In Section 5.4.1 it was shown that η_{j1}, η_{j2}, $|\Delta\eta_{j1j2}|$ and $\eta_{j1}.\eta_{j2}$ appeared to be powerful variables in distinguishing the VBF signal from the data background. It was argued that because variables $|\Delta\eta_{j1j2}|$ and $\eta_{j1}.\eta_{j2}$, are correlated with each other and η_{j1} and η_{j2} that it would only be necessary to use η_{j1} and η_{j2} and the BDT would be able to internally determine $|\Delta\eta_{j1j2}|$ and $\eta_{j1}.\eta_{j2}$. This hypothesis will be put to the test in this section.

Seven η dependent BDT classifiers have been trained. The variables chosen were M_{jj}, $p_{T,j1}$, $p_{T,j2}$ and $|\vec{p}_{T,\gamma\gamma}+\vec{p}_{T,jj}|$ that were decided upon in Section 5.4.1, except η_{j1} and η_{j2} have been removed and replaced with various combinations of η_{j1}, η_{j2}, $|\Delta\eta_{j1j2}|$ and $\eta_{j1}.\eta_{j2}$. For each BDT the significance, contamination and amount of signal was calculated at 100 individual working points. The relationships between significance, contamination and amount of signal have been

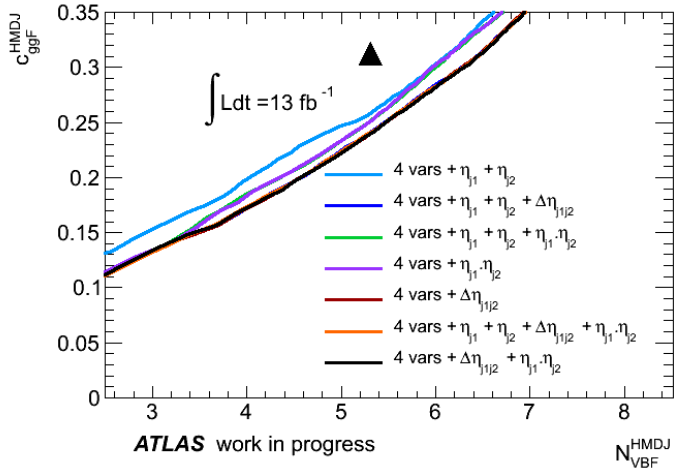
demonstrated in the plots in Figure 5.15. Again the working point for the nominal cut-based working is shown by the black triangle to gauge the amount of improvement that is gained. All seven BDTs give similar similar results. However the BDTs have marked differences when comparing contamination and signal efficiency. All BDTs have a noticeable improvement relative to the cut-based analysis but all the BDTs containing the $|\Delta\eta_{jj}|$ variable, yield a much lower gluon-gluon fusion contamination for a given VBF signal efficiency. The BDT with variables $|\vec{p_{T\gamma\gamma}} + \vec{p_{T,jj}}|$, M_{jj}, $p_{T,j1}$, $p_{T,j2}$ and $|\Delta\eta_{jj}|$ was chosen as this has the fewest variables whilst retaining an equally good performance.

Weak Variables

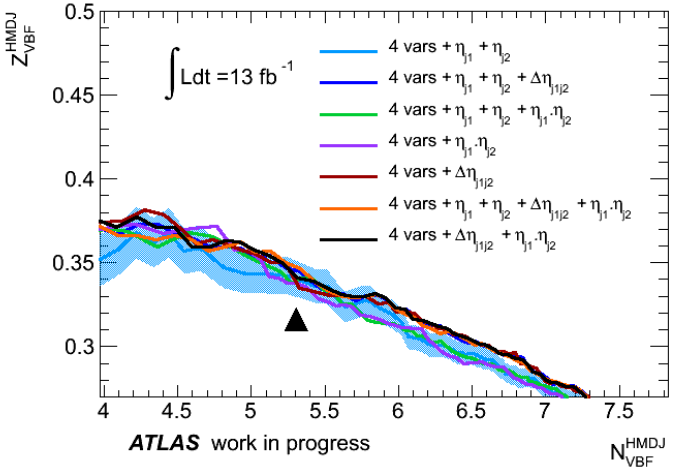
In order to investigate whether adding weak discriminating variables can degrade the performance of the classifier, the azimuthal angle of the leading tag jet (ϕ_{j1}) was included in the training. The distribution of ϕ_{j1} is uniform in both signal and background and therefore has no discriminating power at all. When ϕ_{j1} was added to the classifier in addition to $|\vec{p_{T\gamma\gamma}} + \vec{p_{T,jj}}|$, M_{jj}, $p_{T,j1}$, $p_{T,j2}$ and $|\Delta\eta_{jj}|$, there was shown to be no gain in performance or increase in c_{ggF}^{HMDJ} in both testing and training, as demonstrated in Figure 5.16.



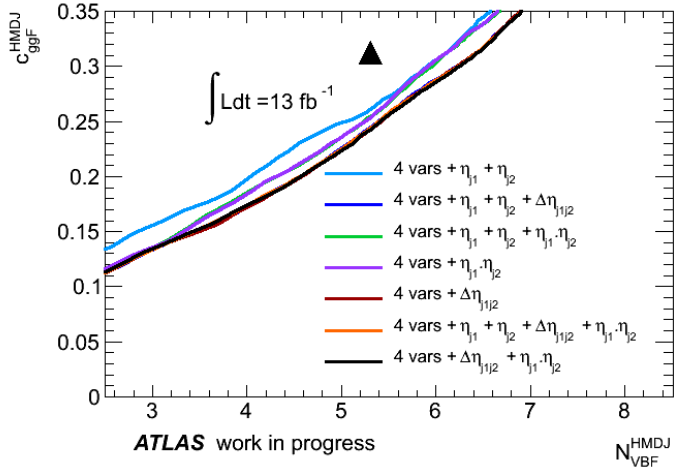
(a) Z_{VBF}^{HMDJ} vs N_{VBF}^{HMDJ} training sample



(b) c_{ggF}^{HMDJ} vs N_{VBF}^{HMDJ} training sample

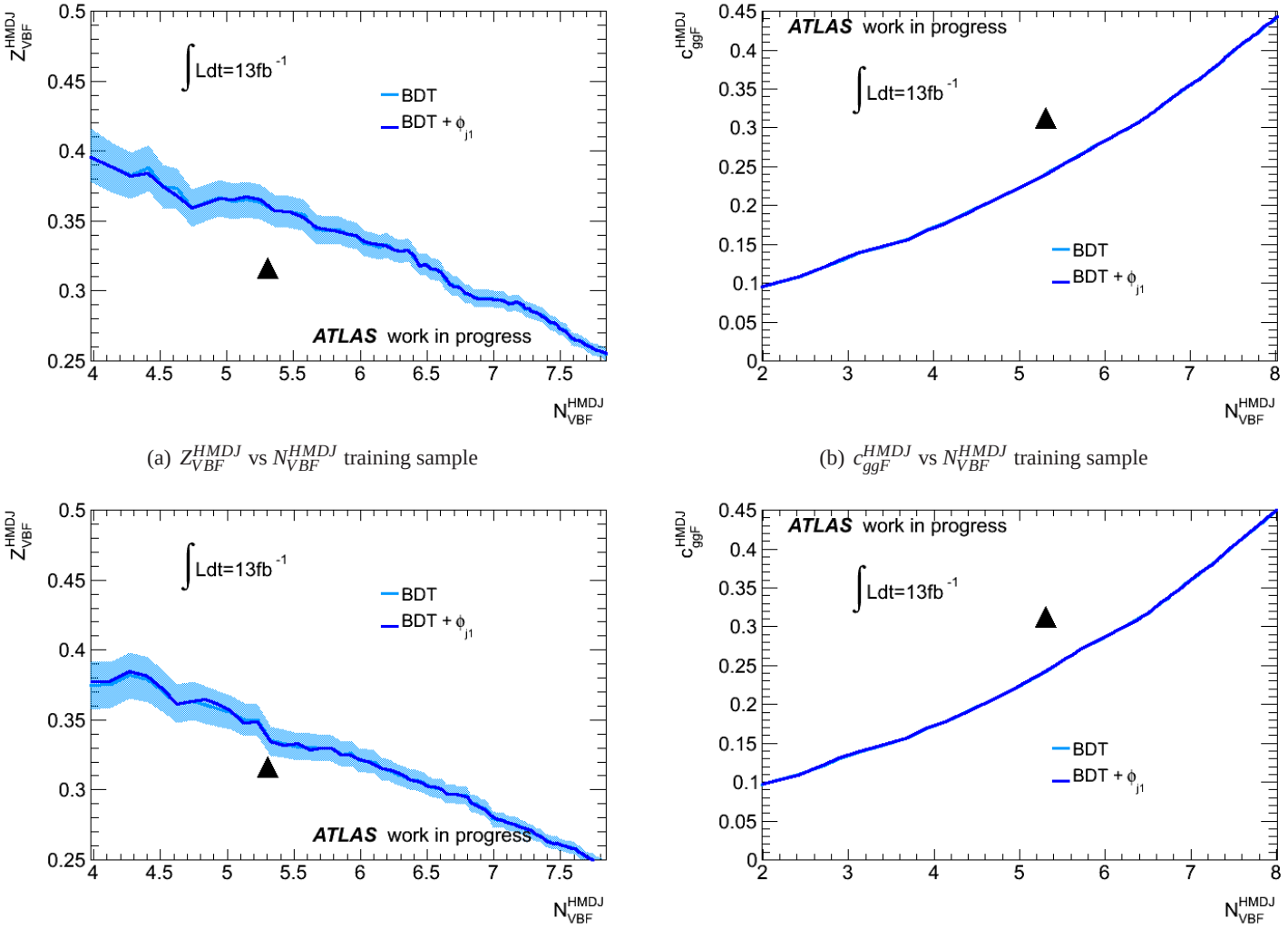


(c) Z_{VBF}^{HMDJ} vs N_{VBF}^{HMDJ} testing sample



(d) c_{ggF}^{HMDJ} vs N_{VBF}^{HMDJ} testing sample

Figure 5.15: Effects of tag jet η variables on the performance of the classifier.



(a) Z_{VBF}^{HMDJ} vs N_{VBF}^{HMDJ} training sample

(b) c_{ggF}^{HMDJ} vs N_{VBF}^{HMDJ} training sample

(c) Z_{VBF}^{HMDJ} vs N_{VBF}^{HMDJ} testing sample

(d) c_{ggF}^{HMDJ} vs N_{VBF}^{HMDJ} testing sample

Figure 5.16: Effects of a weak variable (ϕ_{j1}) on the performance of the classifier.

5.5 Discussion

In this chapter optimisation of the HMDJ category was investigated by either using the cut-based selection by loosening the p_T thresholds on the tag jets or using a BDT classifier. Loosening the p_T thresholds on the tag jets increased the VBF selection efficiency in the HMDJ category but showed little improvement in terms of signal significance. The BDT classifier demonstrated that a higher signal significance could be achieved with VBF selection efficiency comparable to that of the cut-based selection. In addition, the BDT is able to reduce the gluon-gluon fusion signal contamination in the HMDJ category with respect to the cut-based selection. It was therefore decided that a BDT will be used. The BDT that appears to be best in terms of VBF signal efficiency, VBF signal significance and gluon-gluon fusion contamination is the BDT formed of the variables $|\vec{p_{T\gamma\gamma}} + \vec{p_{T,jj}}|$, M_{jj}, $p_{T,j1}$, $p_{T,j2}$ and $|\Delta\eta_{jj}|$. In later chapters a relative measurement of the VBF and gluon-gluon fusion cross sections will also be considered and further improvement to the BDT will be investigated.

The internal configuration of the BDT was also checked to make sure the performance is stable. It was shown that any deviation from the recommended values will have little effect on the performance, so the recommended values will be used throughout.

Chapter 6

Background and Signal Estimation for the Measurement of $\mathfrak{R}$

In this chapter the procedure is setup in which to calculate the fraction, $\mathfrak{R}$, of Higgs boson events produced by VBF relative to the amount of Higgs boson events produced by gluon-gluon fusion and VBF:

$$\mathfrak{R} = \frac{\sigma_{VBF}}{\sigma_{VBF} + \sigma_{ggF}} \tag{6.1}$$

where σ_{ggF} and σ_{VBF} are the cross section of gluon-gluon fusion and VBF respectively. The Standard Model prediction is 0.075 [21]. This will be an extra indicator to test the Standard Model prediction on the newly discovered Higgs-like boson. $\mathfrak{R}$ is of particular interest because gluon-gluon fusion and VBF are the two highest rate production mechanisms of the Higgs boson and will also provide information on the Higgs couplings. There are fermion couplings in the gluon-gluon fusion diagram, and weak boson couplings in the in the VBF diagram. In Section 6.1 the methodology to extract $\mathfrak{R}$ from the data is given. A key aspect of this methodology is determining the amount of background in the signal region; this is described in Section 6.2. Various orders of Bernstein polynomial functions have been investigated as potential candidates to describe the background. In Section 6.3 the potential systematic errors associated with the background modelling are investigated. In Section 6.4 $\mathfrak{R}$ is calculated from pseudodata, to investigate the convergence of the measurement method and the amount of statistical uncertainty.

6.1 Measurement of $\mathfrak{R}$

The amount of signal events in the HMDJ categories and the GGFE category are used to infer $\mathfrak{R}$. From Table 5.1 in Section 5.2 it is safe to assume that the leading contribution of signal events in the HMDJ and the GGFE categories are overhelmingly VBF and gluon-gluon fusion and not other signal sources. (99.4% in the HMDJ category and 96.4% in the GGFE category.) Therefore the total number of signal events of each category (c) is approximated as

$$N^{SR,c} = (\sigma_{ggF}\varepsilon_{ggF}^{SR,c} + \sigma_{VBF}\varepsilon_{VBF}^{SR,c})\mathcal{L}\mathrm{Br}(H \to \gamma\gamma) \tag{6.2}$$

where $\sigma_{ggF(VBF)}$ is the ggF (VBF) cross section, $\mathcal{L}$ is the integrated luminosity, and $\varepsilon_{ggF(VBF)}^{SR,c}$ is the ggF (VBF) signal selection efficiency in category c, which is determined from the signal Monte Carlo samples used in Chapter 5, generated with $m_H = 125\,\mathrm{GeV}$. The amount of signal, $N_s^{SR,c}$ in the signal region is calculated by subtracting the estimated background from the total number of events observed in the signal region, $N^{SR,c}$.

$$N_s^{SR,c} = N^{SR,c} - N_{bkg}^{SR,c} \tag{6.3}$$

If the number of signal events extracted in the signal region of category c, $N^{SR,c}$ is known Equation 6.2 can be inverted to determine the cross sections, σ. Using both the GGFE and the HMDJ categories, one can create at set of simultaneous equations from which to obtain the cross section of each process:

$$\begin{pmatrix} \sigma_{ggF} \\ \sigma_{VBF} \end{pmatrix} \mathrm{Br}(H \to \gamma\gamma) = \\ \frac{1}{\mathcal{L}} \frac{1}{\varepsilon_{ggF}^{SR,HMDJ}\varepsilon_{VBF}^{SR,GGFE} - \varepsilon_{VBF}^{SR,HMDJ}\varepsilon_{ggF}^{SR,GGFE}} \begin{pmatrix} \varepsilon_{VBF}^{SR,GGFE} & -\varepsilon_{VBF}^{SR,HMDJ} \\ -\varepsilon_{ggF}^{SR,GGFE} & \varepsilon_{ggF}^{SR,HMDJ} \end{pmatrix} \begin{pmatrix} N^{SR,HMDJ} \\ N^{SR,GGFE} \end{pmatrix} \tag{6.4}$$

But $\mathfrak{R}$ is the desired result, therefore this is

$$\begin{aligned}\mathfrak{R} &= \frac{\sigma_{VBF}}{\sigma_{VBF}+\sigma_{ggF}} \\ &= \frac{\varepsilon_{VBF}^{SR,GGFE} N_s^{SR,HMDJ} - \varepsilon_{VBF}^{SR,HMDJ} N_s^{SR,GGFE}}{(\varepsilon_{VBF}^{SR,GGFE} N_s^{SR,HMDJ} - \varepsilon_{VBF}^{SR,HMDJ} N_s^{SR,GGFE}) + (\varepsilon_{ggF}^{SR,HMDJ} N_s^{SR,GGFE} - \varepsilon_{ggF}^{SR,GGFE} N_s^{SR,HMDJ})}\end{aligned} \tag{6.5}$$

Note that this is independent of both the integrated luminosity and the $H \to \gamma\gamma$ branching ratio, so these two factors will not contribute to the uncertainty. In order to measure $N^{SR,c}$ the background has to be estimated in the signal region; the procedure to do this is described in the next section.

6.2 Background Estimation in the Signal Region

6.2.1 Background Models

The amount of background is estimated by fitting a function $f(m_{\gamma\gamma};\vec{\theta})$ to the data which is binned every 1 GeV in both of the sidebands ($100 < m_{\gamma\gamma} < 120$ GeV) and ($130 < m_{\gamma\gamma} < 160$ GeV) as it is assumed the sidebands are background. The signal region is blinded so as not to let the signal events bias the position of the fit. The function is integrated with respect to $m_{\gamma\gamma}$ in the signal region to obtain the amount of background in the signal region:

$$N_{bkg}^{SR,c} = \int_{120\,\text{GeV}}^{130\,\text{GeV}} f(m_{\gamma\gamma};\vec{\theta})\,dm_{\gamma\gamma} \tag{6.6}$$

$\vec{\theta}$ is a set of k adjustable parameters θ_i $i = 1,\ldots,k$. As the data points are Poisson distributed the fitting was performed using an extended maximum likelihood for binned data. The log-likelihood is given by

$$-\ln L(\vec{\theta}) = -\sum_{b=1}^{N_{bins}} n_b \ln f_b(\vec{\theta}) - f_b(\vec{\theta}) \tag{6.7}$$

where N_{bins} is the total number of bins, n_b is the number of data events in bin b and $f_b(\vec{\theta})$ is the integral of the fit function between the bin boundaries. As all bins have the same width, h:

$$f_b(\vec{\theta}) = \int_{100\,\text{GeV}+h(b-1)}^{100\,\text{GeV}+hb} f(m_{\gamma\gamma};\vec{\theta})\,dm_{\gamma\gamma} \tag{6.8}$$

The ROOT TMinuit tool [60] was used to maximise the likelihood function such that

$$\frac{\partial L}{\partial \theta_i} = 0 \tag{6.9}$$

at which point, the adjustable parameters tend to their true values.

The amount of background in the signal region, is determined as a function of the estimated parameters $\hat{\vec{\theta}}$ after the log likelihood of $f(m_{\gamma\gamma};\vec{\theta})$ is maximised:

$$\hat{I} = I(\hat{\vec{\theta}}) = \int_{120\,\mathrm{GeV}}^{130\,\mathrm{GeV}} f(m_{\gamma\gamma};\hat{\vec{\theta}})\,dm_{\gamma\gamma} = N_{bkg}^{SR,c} \tag{6.10}$$

The error associated with the background estimation is determined through error propagation [61]:

$$\delta N_{bkg}^{SR,c} = \sum_{i=1}^{k}\sum_{j=1}^{k} \frac{\partial \hat{I}}{\partial \hat{\theta}_i}\frac{\partial \hat{I}}{\partial \hat{\theta}_j} V_{ij} \tag{6.11}$$

where k is the number of adjustable parameter and V_{ij} is the covariance matrix associated with the fit, obtained from TMinuit. The derivatives are obtained using finite difference approximation.

$$\frac{\partial \hat{I}}{\partial \hat{\theta}_k} \approx \frac{I(\hat{\theta}_k + \Delta\theta_k) - I(\hat{\theta}_k - \Delta\theta_k)}{2\Delta\theta_k} \tag{6.12}$$

The value used for $\Delta\theta_k$ is 10% of the fit error on θ_k and is also obtained from TMinuit. $\Delta\theta_k$ is chosen so that it is not too small, so to avoid numerical errors and not too large so that non-linearities in $I(\hat{\vec{\theta}})$ are avoided. [61]

6.2.2 Choice of Model

The function $f(m_{\gamma\gamma};\vec{\theta})$ is a priori, therefore one must take an educated guess of the type of function and the number of adjustable parameters. A log likelihood ratio is used to test the 'goodness' of fit [62]:

$$q_{\vec{\nu}} = -2\ln\frac{L(\vec{\nu})}{L(\hat{\vec{\nu}})} = 2\sum_{b=1}^{N_{bins}} n_b \ln\frac{n_b}{\nu_b} + \nu_b - n_b \tag{6.13}$$

where ν_b is the number of events in bin b associated with the function $f(m_{\gamma\gamma};\vec{\theta})$, $L(\vec{\nu})$ is the likelihood, associated with the function and $L(\hat{\vec{\nu}})$ is the maximum likelihood estimator. $q_{\vec{\nu}}$ will have a higher value for a function that fits the data well compared to that of a function which is a

poor fit to the data.

The goodness of fit can be increased by adding more adjustable parameters. For some functions, if enough parameters were added, the function would fit through each data point. However a model of the background this complex is unlikely, given that $\sim 1\sigma$ fluctuations of the data points would be expected above and below the function. A procedure was formalised to know when to stop adding parameters. It was chosen to adopt the procedure described in [62], which uses a set of "nested functions", eg polynomials of increasing order. These functions are parametrised with parameter set $\vec{\theta}$.

A "p-value" [63], defined as

$$p = \int_{q_{\vec{\nu}}}^{\infty} \frac{1}{2^{N_{bins}/2}\Gamma(N_{bins}/2)} z^{N_{bins}/2-1} e^{-z/2} dz \tag{6.14}$$

can be calculated for increasing orders of polynomial, i.e. adding more parameters. A poor fit (with k adjustable parameters) will correspond to an extremely low p-value meaning there is a small probability of observing the data result assuming the fit hypothesis is true. When the p-value > 0.2 [62], this function should adequately model the background. Ideally, the fit with the highest p-value will be chosen but as the function gets more complex the statistical error on the fit will also increase. An alternative choice, is to calculate another test statistic, which quantifies the improvement a more general function with k+1 parameters has compared with a function of k parameters. This statistic is related to the ratio of maximum likelihoods of the two models

$$q_{k,k+1} = -2\ln \frac{L(\hat{\vec{\theta}}^{(k)})}{L(\hat{\vec{\theta}}^{(k+1)})} \tag{6.15}$$

An associated p-value can be calculated for this test statistic. When the p-value > 0.2 the more general function of k+1 parameters can be rejected with enough confidence.

6.2.3 Bernstein Polynomials

As a benchmark, the fitting procedure was carried out on the HMDJ and the GGFE categories in the nominal cut-based categorisation. Later on, this same procedure will be applied for the BDT categorisation. Bernstein polynomials (BP) are chosen to fit the sidebands, which are constructed

from base polynomials of the form

$$B(x,\vec{\theta}) = \sum_{k=0}^{n} \theta_k \binom{n}{k} x^k (1-x)^{n-k} \tag{6.16}$$

This type of function has been chosen, so as to keep consistency with the background modelling performed by the ATLAS collaboration, which also use Bernstein polynomials for certain categories. Bernstein polynomials, also have the advantage that they are always positive and can be made more general by increasing the order of the polynomial, which is the same as adding parameters as described previously. Corresponding $q_{\vec{v}}$ and p values are shown in Table 6.1 for Bernstein polynomials functions from zeroth order to fifth order.

Order	$q_{\vec{v}}$	$p(q_{\vec{v}})$	$q_{k,k+1}$	$p(q_{k,k+1})$	$N_{bkg}^{SR,HMDJ}$	$\delta N_{bkg}^{SR,HMDJ}$	DOF
0^{th}	52.8385	0.328145	20.002	$\sim$0	54.7997	3.31058	49
1^{st}	32.8358	0.953454	0.0280	0.8941	59.5397	3.74956	48
2^{nd}	32.8078	0.942193	0.0790	0.7859	60.1643	7.14322	47
3^{rd}	32.7287	0.929683	0.9460	0.3310	61.2581	6.63137	46
4^{th}	31.7827	0.931481	0.0330	0.8781	58.7203	6.07484	45
5^{th}	31.7494	0.916005	-	-	58.0913	9.88806	44

Table 6.1: $q_{\vec{v}}$ and $p(q_{\vec{v}})$ values are shown to demonstrate the goodness of fit of Bernstein polynomials of various orders to the data sidebands in the HMDJ category. The values of $q_{k,k+1}$ and $p(q_{k,k+1})$ are shown to test for significant gain from one order to another. The expected background and associated error in the signal region for each fit are also shown.

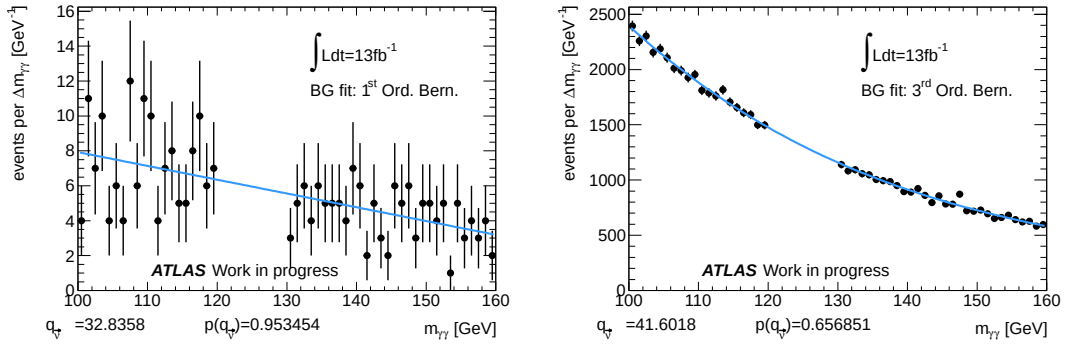
Order	$q_{\vec{v}}$	$p(q_{\vec{v}})$	$q_{k,k+1}$	$p(q_{k,k+1})$	$N_b^{SR,HMDJ}$	$\delta N_b^{SR,HMDJ}$	DOF
0^{th}	12337.4	$\sim$0	11901.0	$\sim$0	12593.8	50.1873	49
1^{st}	436.381	$\sim$0	389.622	$\sim$0	14263.1	60.2359	48
2^{nd}	46.1433	0.507965	4.53500	0.03328	13244.5	75.9261	47
3^{rd}	41.6018	0.656851	0.05700	0.82247	13104.8	99.2896	46
4^{th}	41.5446	0.619127	0.63400	0.42628	13096.0	105.454	45
5^{th}	41.5358	0.577805	-	-	13086.2	144.91	44

Table 6.2: $q_{\vec{v}}$ and $p(q_{\vec{v}})$ values are shown to demonstrate the goodness of fit of Bernstein polynomials of various orders to the data sidebands in the GGFE category. The values of $q_{k,k+1}$ and $p(q_{k,k+1})$ are shown to test for significant gain from one order to another. The expected background and associated error in the signal region for each fit are also shown.

Using $p(q_{\vec{v}})$ obtained from all the functions, it is shown that all orders of function adequately describe the HMDJ background and orders 2 to 5 adequately describe the GGFE background. As already stated the statistical uncertainty on the fit increases for the higher orders as the function

becomes more complex. It is therefore best to choose (within reason) the lowest order of Bernstein polynomials. The $q_{k,k+1}$ and $p(q_{k,k+1})$ values are also shown in Table 6.2. It is shown that $p(q_{k,k+1})$ are a acceptable value at the 1^{st} order in the HMDJ category and 3^{rd} order for the GGFE category. Any higher order can be rejected because no significant gain can be obtained by making the functions more complex.

Based on this analysis, the chosen fit functions for the background in the HMDJ and GGFE categories are a 1^{st} order Bernstein polynomial and a 3^{rd} order Bernstein polynomial, respectively (see Figure 6.1)



(a) HMDJ (b) GGFE

Figure 6.1: Fits to the 13 fb^{-1} of data in the sidebands. The choice of function is explained in the text. (a) First order Bernstein polynomial was chosen to fit the HMDJ category and (b) 3^{rd} order Bernstein polynomial was chosen to fit the GGFE category.

6.3 Potential Systematics of the Background Estimation

Recall that only half of the sideband data sample were used to train the BDT. However the whole (inclusive) data sideband sample is used to fit the background. There is a potential, for the events in the sideband training sample to be underestimated due to possible overtraining of the BDT (i.e. an overtrained BDT could be more efficient than average at rejecting background events that were originally in the training). This is illustrated in Figure 6.2 in a schematic showing the $m_{\gamma\gamma}$ plot of the inclusive events categorised by a BDT. It is shown that there is a higher proportion of testing events in the sidebands.

The other half of the data sidebands (testing sample) was used to check the affect that this has on the background estimation. This was done by fitting to the testing sideband events and the

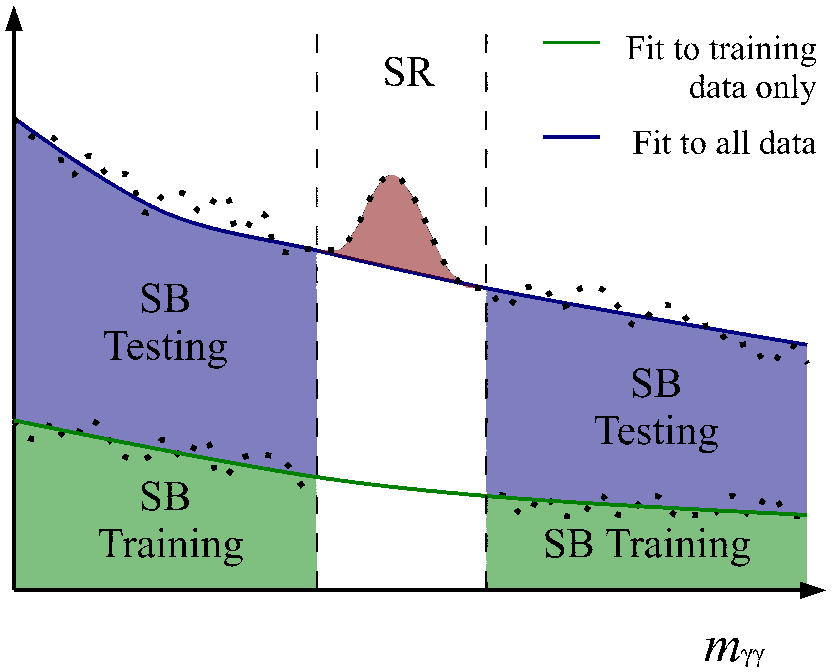


Figure 6.2: Schematic of showing the contribution of data sidebands events in the HMDJ or GGFE category that were used to train the BDT (Green), compared with that of the amount of data sidebands events in the HMDJ or GGFE category that were used to test the BDT(Blue). The relative contributions are exaggerated for the purpose of illustration. The estimated signal (red) is extracted by subtracting the background fitted in the signal region.

training sideband events separately for both the HMDJ and GGFE categories and then scaling to $13\,\text{fb}^{-1}$ (i.e. a factor of 2). This was straightforward for the fit to the HMDJ category but the GGFE category required slight modification to the event selection. Recall that the data in GGFE category has contributions from three subcategories “zero jet”, “one jet” and events that have two jets but are rejected by the BDT (“two jet fail”). Since half of the original statistics from the “two jet fail” subcategory will contribute to GGFE, half of the statistics have to be removed in the “zero jet” and “one jet” subcategories to ensure the relative proportion of each subcategory is equivalent. This procedure is shown in the flow chart in Figure 6.3, which is a modification of the original events selection flow chart in Figure 5.1 with the modifications just discussed shown in yellow.

Using the BDT from Chapter 5 as an example, a working point was chosen, which yielded the same signal yield as the nominal cut base analysis. Using this working point, the BDT yields 109 data sidebands events in the HMDJ category with the training sample and 123 with the testing sample. Although the quantity of these events are just within statistical uncertainty, there still

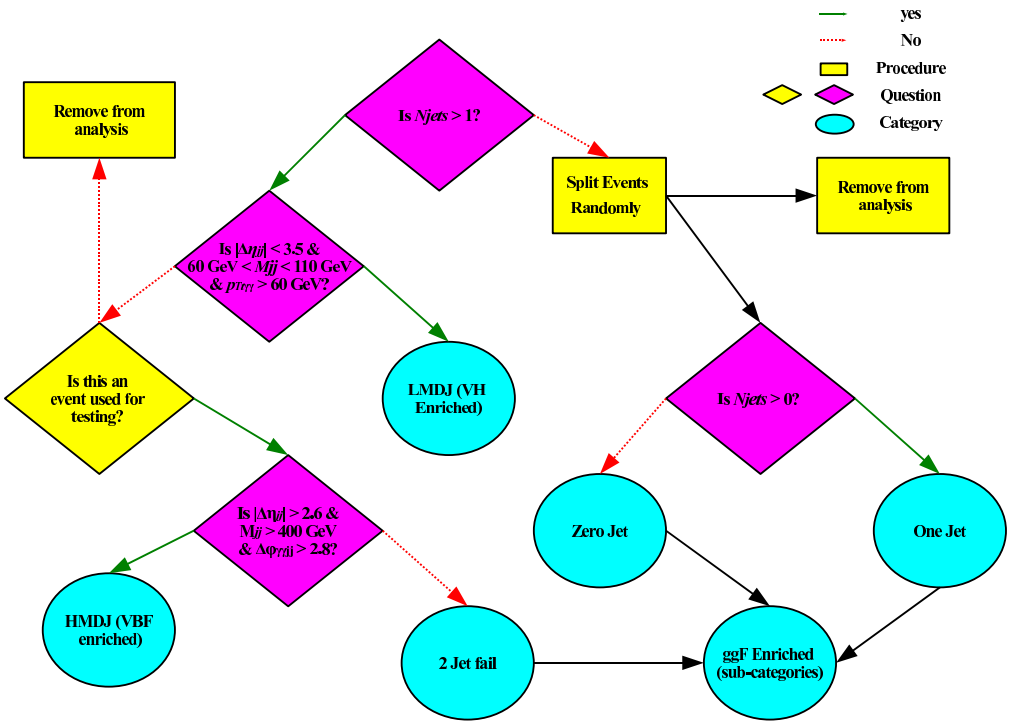


Figure 6.3: Modification to the event selection so that the integrated luminosity of the testing sub-sample of the GGFE category is equivalent to the testing sub-sample of the HMDJ category.

could be a potential bias. The choice of fit function was therefore re-evaluated separately in the testing and training samples, each with half the original statistics. The background estimate in the signal region and the statistical error was then scaled by a factor of two to restore to $\mathcal{L} = 13\,\text{fb}^{-1}$. The background estimates are shown in Table 6.3. The background estimates in the signal region are very similar. The training and the tesing subsamples agree within statistical error and therefore a bias is ruled out from the inclusive estimate.

Another potential systematic was investigated that could have arose due to the nature of the Bernstein polynomials functions. The concern was that for higher orders of Bernstein polynomials, the function could become less monotonic, which could create a bias in the signal region. In particular the 3^{rd} order Bernstein polynomials has a base function which has a maximum in the centre of the signal region, (see Figure 6.4). Given that no data points in the signal region are included in the fit, this could enhance the total Bernstein polynomials function in the signal region and over-estimate the background in the signal region. By comparing fits to the background

Category	Testing Events			Training Events			Inclusive
	$N_{bkg}^{SR,c}$	$\delta N_{bkg}^{SR,c}$	Scaled to 13fb^{-1}	$N_{bkg}^{SR,c}$	$\delta N_{bkg}^{SR,c}$	Scaled to 13fb^{-1}	
HMDJ	26.496	2.48642	52.992	23.865	2.38386	47.730	50.337
GGFE	6530.87	70.224	13061.7	6530.30	70.275	13060.6	13111.2

Table 6.3: The expected amount of background in the signal region of both the HMDJ category and the GGFE category. The expectations are determined from fitting to the sidebands in the testing and training samples. As both testing sample and training samples are half the 13fb^{-1}, the results were scaled by a factor 2 and compared with the inclusive fit, which is where the testing and training samples are combined together.

with Bernstein polynomials of different orders (k =2,3 or 4), it has been checked that the relative magnitudes of the basis polynomials are consistant in all the fits. It was therefore considered acceptable to proceed with Bernstein polynomials. Alternative orders of Bernstein polynomials were used to calculate systematic uncertainties as will be described in Chapter 8.

6.4 Estimating $\mathfrak{R}$ and its Uncertainty using Pseudodata

Now that a method of determining N_s^{SR} has been established, $\mathfrak{R}$ can be determined. However in order to do this, the amount of data in the signal region will have to be revealed. Since a working point has not yet been decided upon, the signal region cannot be unblinded, as this could potentially bias the decision. However the likely value of $\mathfrak{R}$ that will be measured if the signal region were to be unblinded can be investigated using MC-based 'pseudodata'.

1,000,000 pseudodata samples (toy experiments) were generated, each toy experiment representing 13 fb^{-1} of data, where a value of $\mathfrak{R}$ is calculated for each. In order to calculate $\mathfrak{R}$ in each toy experiment, the expected number of Higgs boson events in the signal region of each category, c, $N_{s,toy}^{SR,c}$, has to be obtained by subtracting the expected number of background events in the signal region, $N_{bkg,toy}^{SR,c}$ from the total number of events $N_{toy}^{SR,c}$

$$N_{s,toy}^{SR,c} = N_{toy}^{SR,c} - N_{bkg,toy}^{SR,c} \tag{6.17}$$

$N_{bkg,toy}^{SR,c}$ and $N_{toy}^{SR,c}$ are determined from random number generation in each toy experiment. $N_{bkg,toy}^{SR,c}$ is determined from a Gaussian random number generator using $N_{bkg}^{SR,c}$ as the mean and the statistical

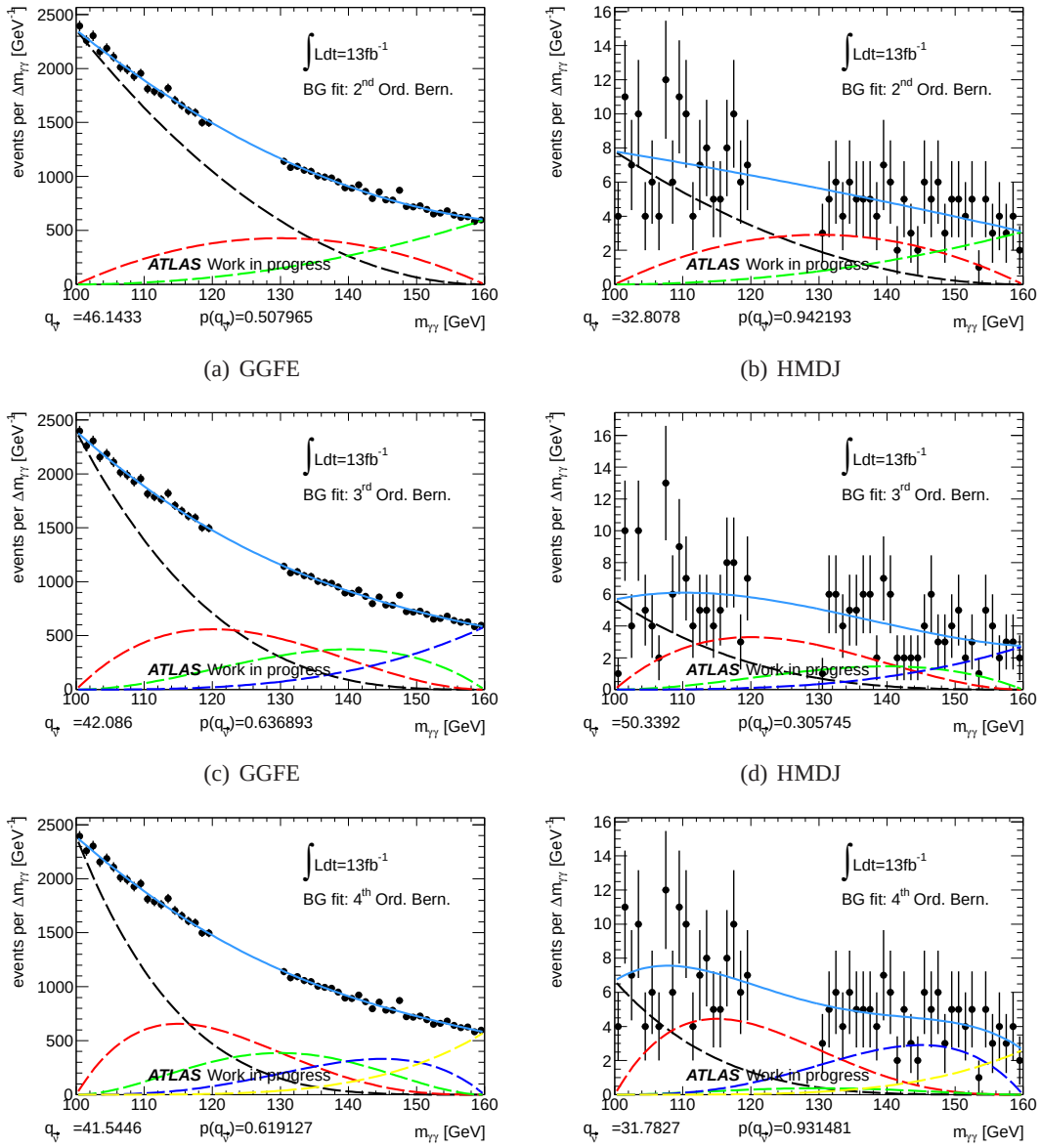


Figure 6.4: Base components for different orders of Bernstein polynomials that were fitted to the sidebands of the HMDJ and GGFE categories.

error obtained by the fit $\delta N_{bkg}^{SR,c}$ as the spread of the Gaussian

$$N_{bkg,toy}^{SR,c} = Gaus(\nu = N_{bkg}^{SR,c}, \sigma = \delta N_{bkg}^{SR,c}) \tag{6.18}$$

$N_{toy}^{SR,c}$ is determined from a Poisson random number generator

$$N_{toy}^{SR,c} = Pois(\nu = N_{VBF}^{SR,c} + N_{ggF}^{SR,c} + N_{bkg}^{SR,c}) \tag{6.19}$$

using a mean value, ν, which is the sum of the VBF signal, $N_{VBF}^{SR,c}$ and the gluon-gluon fusion contribution, $N_{ggF}^{SR,c}$ from the SM prediction.

$$N_{ggF(VBF)}^{SR,c} = \varepsilon_{ggF(VBF)}^{SR,c} \sigma_{ggF(VBF)}^{SM} \mathcal{L} \mathrm{Br}(H \to \gamma\gamma) \tag{6.20}$$

where the efficiency, $\varepsilon_{ggF(VBF)}^{SR,c}$ is obtained from MC and the cross section $\sigma_{ggF(VBF)}^{SM}$ is the SM hypothesis but, in principle, alternative hypotheses can also be investigated.

The first test was to demonstrate that the value of $\mathfrak{R}$ obtained would be consistent with the true value of $\mathfrak{R}$ under the SM hypothesis and 4 alternative hypotheses. Five samples of 1,000,000 toy experiments were generated for five different cross section scenarios:

1. Assume SM σ_{ggF} and σ_{VBF} cross sections and SM $H \to \gamma\gamma$ branching ratio;
2. Same as 1. except $\sigma_{ggF} \to \sigma_{ggF} \times 2$;
3. Same as 1. except $\sigma_{ggF} \to \sigma_{ggF}/2$;
4. Same as 1. except $\sigma_{VBF} \to \sigma_{VBF} \times 2$;
5. Same as 1. except $\sigma_{VBF} \to \sigma_{VBF}/2$.

For the purpose of this demonstration, the background estimates in the signal region were determined using sideband fits to the nominal cut-based HMDJ and GGFE categories.

The values of $\mathfrak{R}$ were binned as shown in Figure 6.5. The distributions in $\mathfrak{R}$ that are shown in Figure 6.5, show an asymmetry, especially for scenarios 3 and 5, where σ_{ggF} and σ_{VBF} are reduced by a factor 2.

This asymmetry can be explained by the nature of a general Poisson distribution. When ν is a low value the distribution is asymmetrical. However for large values of ν the distribution

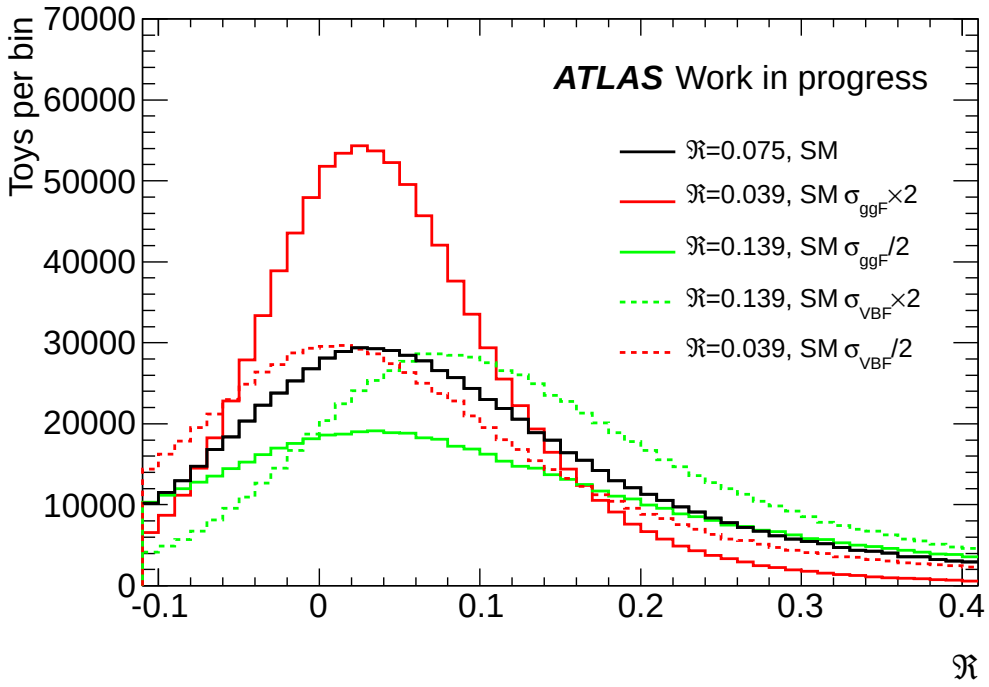
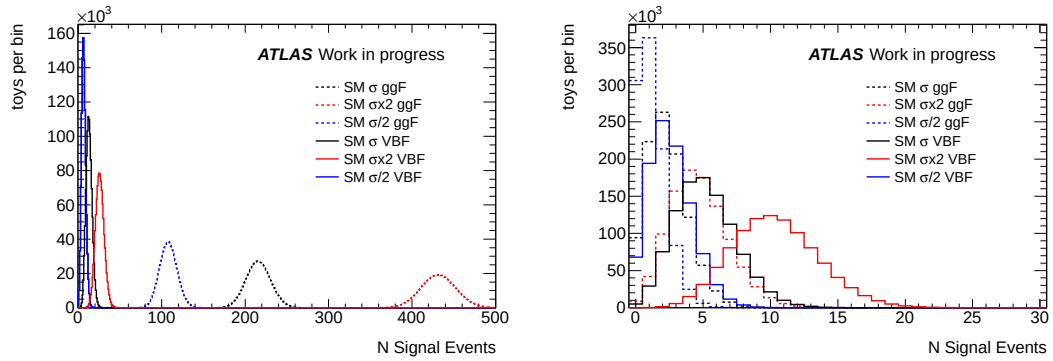


Figure 6.5: Measured $\mathfrak{R}$ with 1,000,000 toys of pseudodata on 5 different cross section hypotheses. The Value of R in each toy is calculated from randomly generated numbers, that are consistent with the expectation of signal and background for each cross section hypothesis.

becomes more Gaussian like. This effect is shown in Figure 6.6. The signal expectation value in the HMDJ category is much smaller than the signal expectation value in the GGFE category where it is shown there is a greater asymmetry in the distributions in Figure 6.6(b) than in Figure 6.6(a). For the $\sigma^{SM}_{ggF(VBF)}/2$ scenarios, the impact of this asymmetry is most noticed in the HMDJ category and is visible in the $\mathfrak{R}$ distribution.

It is expected that with more data, the uncertainty on the $\mathfrak{R}$ measurement will decrease. This can be investigated with additional toy experiments at increased $\mathcal{L}$. The spread of the distributions should decrease and the peak position of $\mathfrak{R}$ should converge to the true value of $\mathfrak{R}$. The five cross section scenarios were regenerated for four alternative values of $\mathcal{L}$: $\mathcal{L} = 50\text{fb}^{-1}$, 100fb^{-1}, 200fb^{-1} and 400fb^{-1}. The background is assumed to scale with the increased $\mathcal{L}$ and the statistical error on the fit is assumed to scale with $\sqrt{\mathcal{L}}$. The results from these alternative scenarios are shown in Figure 6.7 and demonstrate a general trend for the spread of $\mathfrak{R}$ to decrease and the measured values to converge to the true values of $\mathfrak{R}$.

Regardless of the fact that with increasing luminosity the distribution of toy experiments will



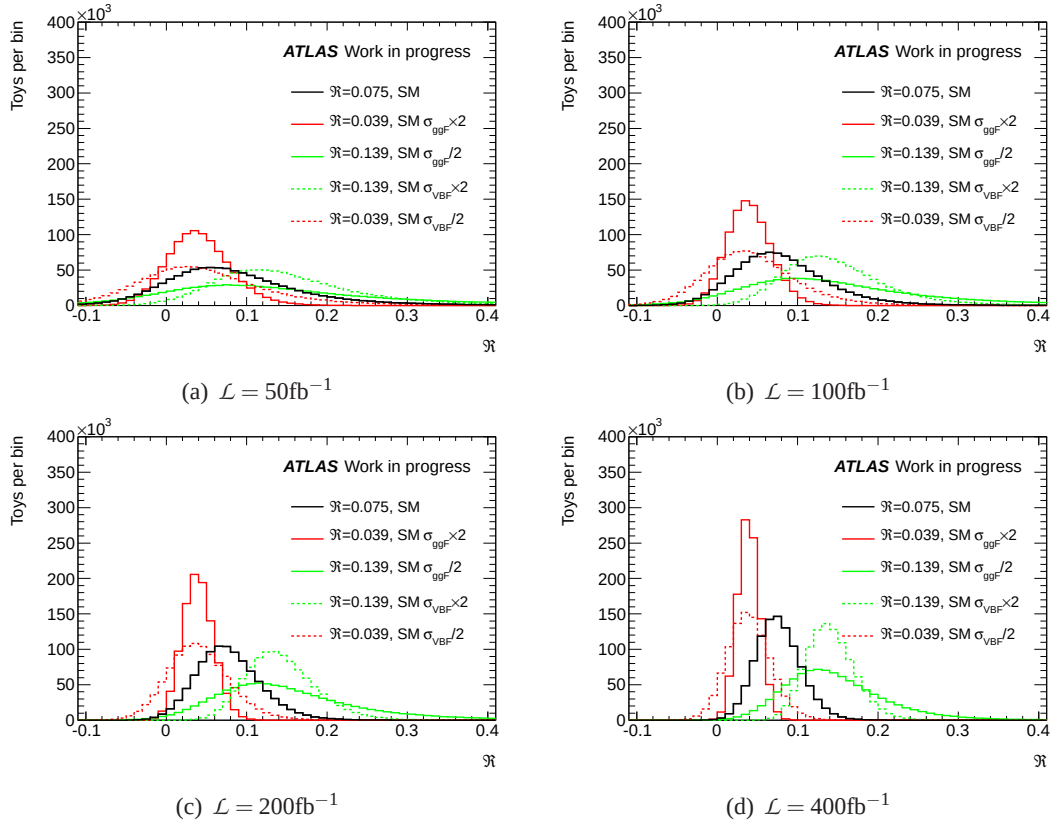
(a) GGFE (b) HMDJ

Figure 6.6: Amount of gluon-gluon fusion signal and VBF signal events for each toy. Each is a randomly generated number, that are consistent with the expection of signal from each cross section hypothesis.

peak at the true value of $\mathfrak{R}$, at 13 fb^{-1} there remains a potential systematic difference between the true value of $\mathfrak{R}$ and the value extracted from the measurement in the data. The associated systematic uncertainty will be quantified in Chapter 8.

6.5 Discussion

It is clear to see that using the nominal cut-based selections will result in a large statistical uncertainty on $\mathfrak{R}$ and more data will be required in order to reduce this uncertainty. The intention is now to optimise the HMDJ category by using a multivariate analysis and working point that will result in a smaller uncertainty and also investigate the effects of gluon-gluon fusion contamination, VBF signal selection efficiency and signal significance have on this measurement.



(a) $\mathcal{L} = 50\text{fb}^{-1}$

(b) $\mathcal{L} = 100\text{fb}^{-1}$

(c) $\mathcal{L} = 200\text{fb}^{-1}$

(d) $\mathcal{L} = 400\text{fb}^{-1}$

Figure 6.7: Measured $\mathfrak{R}$ with 1,000,000 toy experiments of pseudodata for 5 different cross section hypotheses. The value of $\mathfrak{R}$ in each toy experiment is calculated from randomly generated numbers, that are consistent with the expection of signal and background for each cross section hypothesis. This is shown for various data sample sizes; (a) $\mathcal{L} = 50\text{fb}^{-1}$, (b) $\mathcal{L} = 100\text{fb}^{-1}$, (c) $\mathcal{L} = 200\text{fb}^{-1}$ and (d) $\mathcal{L} = 400\text{fb}^{-1}$.

Chapter 7

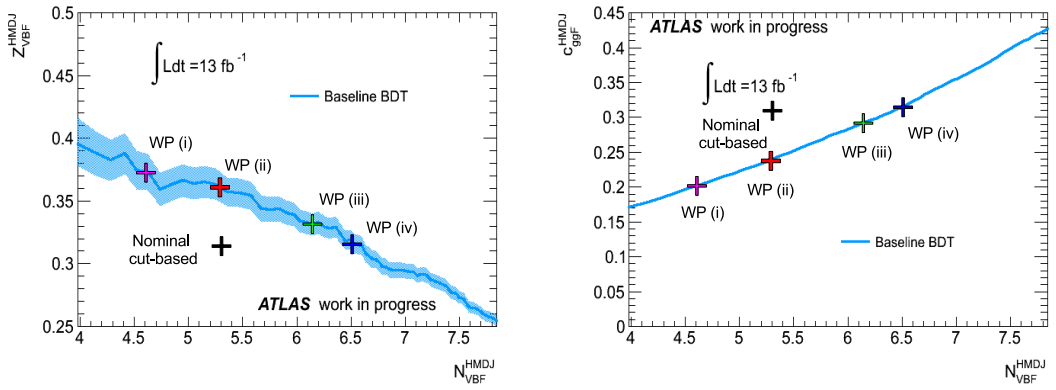
Final Choice of BDT for the Measurement of $\mathfrak{R}$

A baseline BDT classifier based on variables p_{Tj1}, p_{Tj2}, $\Delta\eta_{jj}$, M_{jj} and $|\vec{p}_{T\gamma\gamma}+\vec{p}_{Tjj}|$ was defined in Chapter 5. The BDT was shown to give an improvement on the signal selection efficiency, significance and reduced the gluon-gluon fusion contamination in the HMDJ with respect to the nominal cut-based categorisation. The choice of working point may either favour, disfavour or be a compromise to either one of these metrics. However, the value of $\mathfrak{R}$ is now also desired, which, as demonstrated in Chapter 6, has a large uncertainty because of limited statistics. A working point which minimises the uncertainty on $\mathfrak{R}$ is therefore also desired. An investigation is presented in this chapter to determine which metric best improves the HMDJ selection such that the uncertainty on $\mathfrak{R}$ is minimised. It will be shown that the working points with high signal significance are most likely to achieve this. A high value of Z_{VBF}^{HMDJ} can be obtained by including additional variables in the BDT classifier defined in Chapter 5 or by choosing a working point with lower VBF signal efficiency.

7.1 Choice of Working Point on the Baseline BDT Classifier

Four working points (WP) were chosen from the baseline BDT classifier defined in Chapter 5, which are shown in Figure 7.1. WP(ii) was chosen to have approximately the same signal efficiency expectation as the nominal cut-based selection and WP(iv) was chosen to have approx-

imately the same VBF signal significance as the nominal cut-based selection. Two additional working points, WP(i) and WP(iii), were also chosen so as to better cover the relevant performance region. The working points were chosen on the outcome of BDT using the training sample, so not to bias the choice. It is shown in Figure 7.1(b) that the gluon-gluon fusion signal contamination is reduced for working points that yield lower N_{VBF}^{HMDJ} and higher Z_{VBF}^{HMDJ}. The expected yields predicted with the training sample are shown in Table 7.1.



(a) Z vs N (b) c vs N

Figure 7.1: The coloured crosses represent four possible working points on the 5 variable BDT classifier, which predict N_{VBF}^{HMDJ}, Z_{VBF}^{HMDJ} and c_{ggF}^{HMDJ} from the training in four potential HMDJ categories. The black cross shows the yields for the cut-based analysis. The values of N_{VBF}^{HMDJ} and Z_{VBF}^{HMDJ} are shown in (a) and the values of N_{VBF}^{HMDJ}, and c_{ggF}^{HMDJ} are shown in (b).

WP	N_{VBF}^{HMDJ}	Z_{VBF}^{HMDJ}	c_{ggF}^{HMDJ}
Nominal	5.31	0.316	0.312
i	4.41	0.388	0.191
ii	5.36	0.357	0.243
iii	6.09	0.333	0.288
iv	6.53	0.316	0.317

Table 7.1: The values of of N_{VBF}^{HMDJ}, Z_{VBF}^{HMDJ} and c_{ggF}^{HMDJ} yields from the training sample for the four working points on the 5 variable BDT classifier.

The data and MC signal events were selected using the event selection outlined in Chapter 5. Each working point, defines a potential classifier which replaces the cuts demonstrated by the red diamond box in Figure 5.1. The testing sample of the VBF signal were classified by the BDT and scaled by a factor of two to rescale to 13 fb^{-1} as demonstrated in Chapter 6. Invariant mass distributions were produced with the data for the resultant four HMDJ and four GGFE categories

and functions were fitted to the sidebands using the procedure described in Chapter 6. A 1^{st} order Bernstein polynomial was sufficient to model the background in all HMDJ categories and a 3^{rd} order Bernstein polynomial for all GGFE categories. The predicted background and statistical errors were calculated using the fit function in both categories. The signal efficiencies for gluon-gluon fusion and VBF were determined in the signal region of each of the four resulting HMDJ and GGFE categories. The background and efficiency predictions are shown in Tables 7.2 and 7.3.

WP	$\varepsilon_{ggF}^{SR,HMDJ}$	$\delta\varepsilon_{ggF}^{SR,HMDJ}$	$\varepsilon_{VBF}^{SR,HMDJ}$	$\delta\varepsilon_{VBF}^{SR,HMDJ}$	$N_{bkg}^{SR,HMDJ}$	$\delta N_{bkg}^{SR,HMDJ}$
i	0.00176	0.00003	0.0918	0.0005	28.56	2.62
ii	0.00290	0.00004	0.1109	0.0006	50.34	3.44
iii	0.00416	0.00005	0.1260	0.0006	73.54	4.17
iv	0.00510	0.00006	0.1355	0.0006	91.74	4.08

Table 7.2: Expected signal efficiencies and background in the signal region of potential HMDJ categories defined as explained in the text.

WP	$\varepsilon_{ggF}^{SR,GGFE}$	$\delta\varepsilon_{ggF}^{SR,GGFE}$	$\varepsilon_{VBF}^{SR,GGFE}$	$\delta\varepsilon_{VBF}^{SR,GGFE}$	$N_{bkg}^{SR,GGFE}$	$\delta N_{bkg}^{SR,HMDJ}$
i	0.3755	0.0004	0.2939	0.0008	13135.3	99.4
ii	0.3744	0.0004	0.2748	0.0008	13111.2	99.3
iii	0.3731	0.0004	0.2597	0.0008	13085.1	99.3
iv	0.3722	0.0004	0.2502	0.0008	13066.5	99.2

Table 7.3: Expected signal efficiencies and background in the signal region of potential GGFE categories defined by the classification explained in the text.

1,000,000 MC toy experiments were then generated for each working point to determine the likely value of $\mathfrak{R}$ that would be obtained if it were to be measured using the procedure outlined in Chapter 6. The background in the signal region for each toy experiment was determined using a Gaussian random number generator where the mean, ν and the standard deviation, σ were set to N_{bkg}^{SR} and δN_{bkg}^{SR} respectively, these values are shown in Tables 7.2 and 7.3. The amount of signal in the signal region for each toy experiment was determined using a Poisson random number generator where the mean, ν was set to the SM expectation corresponding to the expected signal efficiencies shown in Tables 7.2 and 7.3. The corresponding distributions of $\mathfrak{R}$, for each working point are shown seperately in Figure 7.2. It can be seen that WP(i) has the narrowest distribution and therefore, indicating the statistical uncertainty will be smallest if $\mathfrak{R}$ were to be measured using this working point.

It is desirable to choose a working point which gives the lowest statistical uncertainty on $\mathfrak{R}$.

WP(ii) was chosen to minimise the uncertainty on $\mathfrak{R}$ whilst retaining a VBF selection efficiency comparable to the nominal cut-based selection. WP(i) would seem like the obvious working point to choose. However the expected number of VBF signal events in the HMDJ category is already limited, choosing WP(i) would further decrease it.

Having chosen WP(ii), the VBF signal efficiency, signal significance and gluon-gluon fusion contamination in the HMDJ category were evaluated using the testing samples. A 5.9% improvement, can be gained on signal significance with respect to the cut-based analysis and the gluon-gluon fusion is decreased by 21.7% with respect to the cut-based analysis.

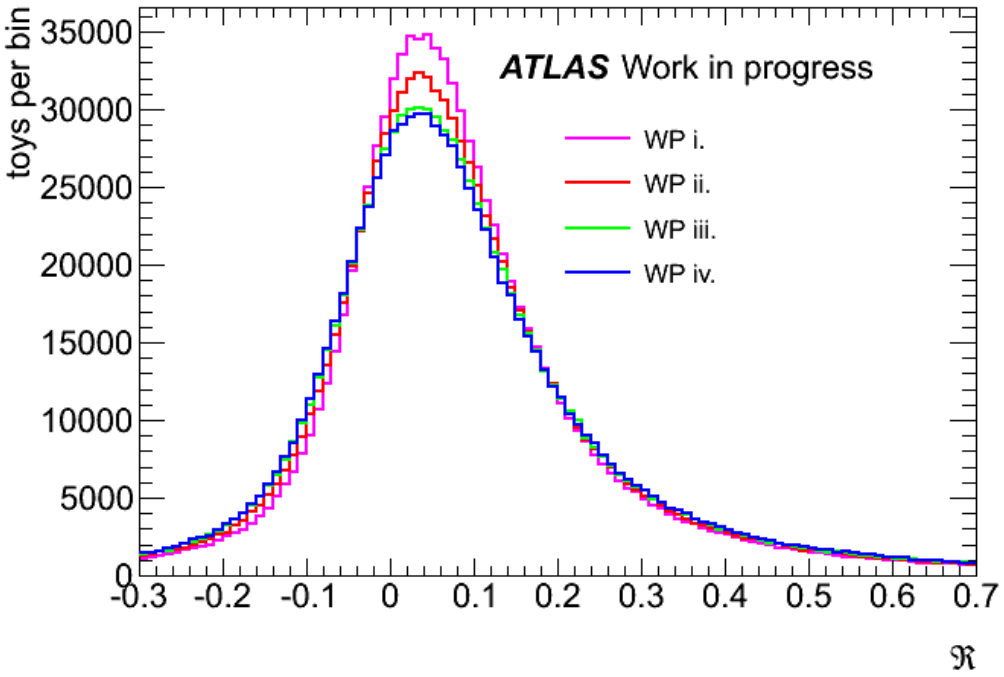


Figure 7.2: Distributions of $\mathfrak{R}$ for the four different working points, obtained using pseudodata. The colours of each histogram correspond to the colours of the crosses on the working points in Figure 7.1.

7.2 Improving the BDT Classifier

The results shown in the previous section would suggest that a working point which yields the largest Z_{VBF}^{HMDJ} is the optimal working point to choose for the best result on $\mathfrak{R}$. However it was concluded that no further gain on signal significance could be gained without loss in VBF signal efficiency with respect to the cut-based categorisation. It is however possible to increase Z_{VBF}^{HMDJ}

further and still retain a VBF signal yield that is comparable to that of the nominal cut-based categorisation, by including additional variables in the classifier. With more variables, the classifier will be able to exploit more information about the signal and background to increase the separation power.

7.2.1 BDT Classifiers with Six Variables

The baseline MVA was constructed out of the Type A variables M_{jj}, $\Delta\eta_{jj}$, p_{Tj1}, p_{Tj2} and $|\vec{p}_{T,\gamma\gamma} + \vec{p}_{T,jj}|$ meaning that these variables provide good VBF signal-background separation and similarities between the background and the gluon-gluon fusion signal. Type B variables are now added to the classifier to investigate if any further improvement can be gained. The signal distributions of the four Type B variables ($\Delta\phi_{\gamma\gamma,jj}$, $p_{T,\gamma 1}$, $\Delta R_{\gamma 1,j1}$ and $p_{T,t\gamma\gamma}$) are similar for that of the gluon-gluon fusion and VBF signals, which has the potential to cause gluon-gluon fusion signal contamination in the HMDJ.

Irrespective of this risk, the four Type B variables were added separately, producing four alternative classifiers (a, b, c and d). The variables that are used for these classifiers are listed in Table 7.4. For each classifier the values of Z_{VBF}^{HMDJ}, N_{VBF}^{HMDJ} and c_{VBF}^{HMDJ} were determined for 100 working points and are shown in Figure 7.3.

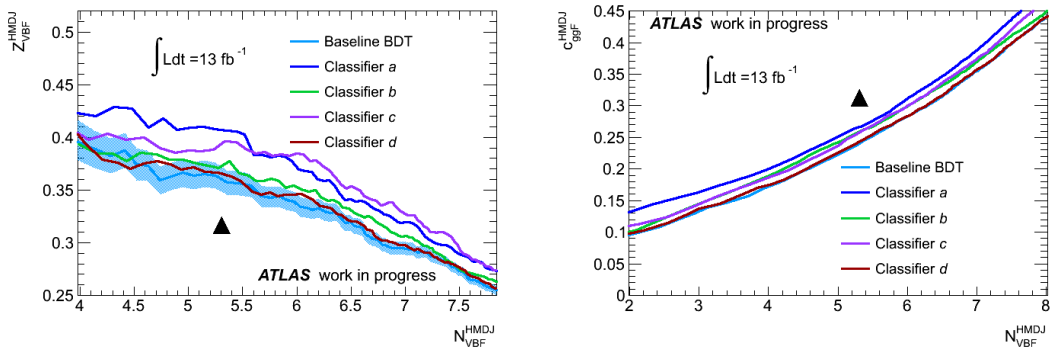
A working point from classifiers *a*, *b*, *c* and *d* can be chosen to yield the same amount of VBF signal as would be obtained from the nominal cut-based categorisation. For all these four working point Z_{VBF}^{HMDJ} is higher with respect to the baseline MVA and the nominal cut-based selection. However, as predicted some classifiers increase the amount of gluon-gluon fusion contamination, which can be seen in classifiers *a*, *b* and *d*. All working points that yield a compatible VBF signal with respect to the nominal cut-based classifier, show an increase in gluon-gluon fusion signal contamination with respect to the 5 variable BDT shown in the previous section. Even so, none of the classifiers, *a*, *b*, *c* or *d* result in gluon-gluon fusion signal contamination that is higher than the nominal cut-based classification, for these given working points.

The KS probabilities are all greater than 0.1, as shown in Table 7.4, indicating that there is no sign of overtraining. Classifier *a*, has shown the highest signal significance, Z_{VBF}^{HMDJ}, so this will now be considered to provide an improved measurement on $\mathfrak{R}$. As classifier *d* has shown little improvement with respect to Z_{VBF}^{HMDJ} or N_{VBF}^{HMDJ} variable $\Delta R_{\gamma 1,j1}$ will henceforth be ignored.

Although classifiers *a*, *b* and *d* have a higher gluon-gluon fusion signal contamination with

Classifier	Input Variables	KS	
		Signal	Background
	Baseline BDT (5 variables)	0.279	0.735
a	Baseline BDT + $p_{T,t\gamma\gamma}$	0.565	0.904
b	Baseline BDT + $\lvert\Delta\phi_{\gamma\gamma,jj}\rvert$	0.321	0.999
c	Baseline BDT + $p_{T,\gamma 1}$	0.357	0.986
d	Baseline BDT + $\Delta R_{\gamma 1,j1}$	0.314	0.985
e	Baseline BDT + $p_{T,t\gamma\gamma}$ + $\Delta\phi_{\gamma\gamma,jj}$	0.638	0.967
f	Baseline BDT + $p_{T,t\gamma\gamma}$ + $p_{T\gamma 1}$	0.575	0.704
g	Baseline BDT + $p_{T,t\gamma\gamma}$ + $\Delta\phi_{\gamma\gamma,jj}$ + $p_{T\gamma 1}$	0.586	0.590

Table 7.4: Variables used to train each classifier. The KS probabilities determind from the testing and training samples is also shown (see Section 5.4.3 for details).



(a) Z_{VBF}^{HMDJ} vs N_{VBF}^{HMDJ} (b) c_{ggF}^{HMDJ} vs N_{VBF}^{HMDJ}

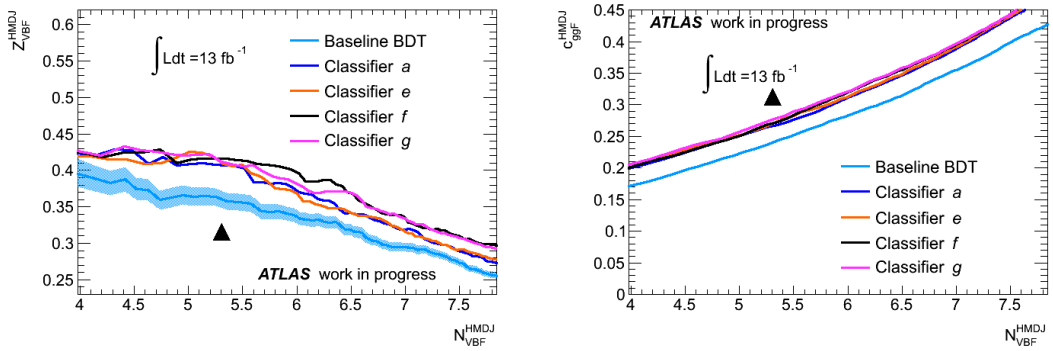
Figure 7.3: Training performance of classifiers which use the variables chosen for the baseline BDT classifier and one of Type B variables. The performance in terms of VBF signal yield and VBF signal significance is shown in (a) and the performance in terms of VBF signal yield and gluon-gluon fusion signal contamination is shown in (b).

respect to the baseline BDT, there is still not as much contamination as there would be if the nominal cut-based selection was used. Out of these three remaining BDTs, classifier *a* was chosen as the best 6 variable BDT, as it gives the best signal significance for a given VBF signal efficiency. Having established this, the VBF signal efficiency, signal significance and gluon-gluon fusion contamination in the HMDJ category were evaluated using the testing samples. A 18.9% improvement, can be gained on signal significance (5.9% for the 5 variable BDT) with respect to the cut-based analysis and the gluon-gluon fusion is decreased by 13.6% with respect to the cut-based analysis (21.7% for the 5 variable BDT).

7.2.2 BDT Classifiers with More Than Six Variables

Investigations were carried out to see if any further improvement could be gained on the performance of the BDT by including two or three additional Type B variables. The combinations of Type B variables are shown in Table 7.4 for four additional classifiers (e, f, g and h). The KS probabilities are shown for each classifier to verify that they are not over-trained. N_{VBF}^{HMDJ}, c_{ggF}^{HMDJ} and Z_{VBF}^{HMDJ} were determined for 100 working points for each classifier and are shown in Figure 7.4. In addition, classifier a and the 5 variable BDT are shown for reference. Where N_{VBF}^{HMDJ} is fixed to that obtained by the nominal cut-based selection slight improvement on Z_{VBF}^{HMDJ} can be gained by training with 7 variables and there is very little improvement by training with 8 variables. The 8 variable classifier, will therefore not be considered any further.

Out of the 7 variable BDTs, classifier f was chosen as this produced the best signal significance without reducing the VBF signal efficiency. Using the testing sample the VBF signal efficiency, signal significance and gluon-gluon fusion contamination in the HMDJ category were evaluated. A 24.0% improvement, can be gained on signal significance with respect to the cut-based analysis (5.9% for the 5 variable BDT) and the gluon-gluon fusion is decreased by 12.0% with respect to the cut-based analysis (21.7% for the 5 variable BDT).



(a) Z_{VBF}^{HMDJ} vs N_{VBF}^{HMDJ} (b) c_{ggF}^{HMDJ} vs N_{VBF}^{HMDJ}

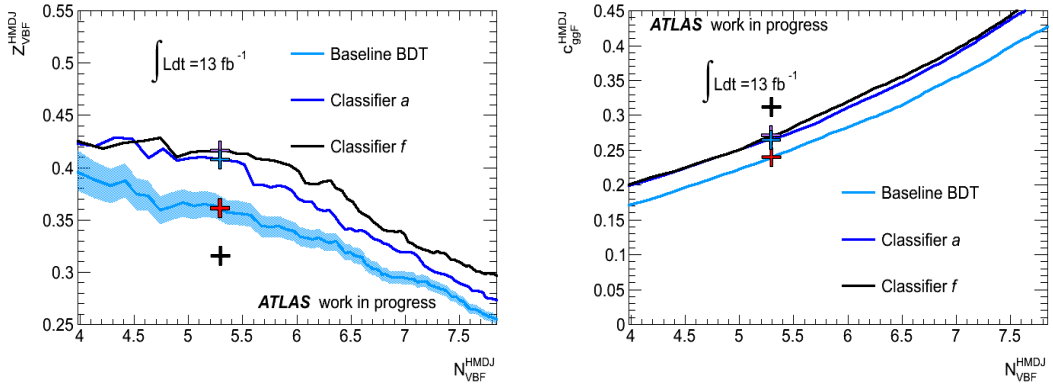
Figure 7.4: Training performance of classifiers which use the variables chosen for the baseline BDT classifier and some additional Type B variables. The performance in terms of VBF signal yield and VBF signal significance, Z_{VBF}^{HMDJ} is shown in (a) and the performance in terms of VBF signal yield and gluon-gluon fusion signal contamination is shown in (b).

7.3 Final Choice of Working Point

A working point will now be decided upon that gives the smallest uncertainty on $\mathfrak{R}$. It is now possible to choose a working point that improves the signal significance with respect to the nominal cut-based analysis for a given signal efficiency which is comparable to the nominal cut-based analysis. A working point was chosen with these criteria for:

- the baseline 5 variable BDT;
- the 6 variable BDT (classifier *a*);
- the 7 variable BDT (classifier *f*).

These are shown in Figure 7.5(a) which follow the Z_{VBF}^{HMDJ} axis from the nominal cut-based categorisation.



(a) Z_{VBF}^{HMDJ} vs N_{VBF}^{HMDJ} (b) c_{ggF}^{HMDJ} vs N_{VBF}^{HMDJ}

Figure 7.5: Working points that have been chosen to yield the same amount of VBF signal with respect to the nominal cut-based analysis but vary in Z_{VBF}^{HMDJ}. The nominal cut-based categorisation of events is shown by the black cross.

WP	N_{VBF}^{HMDJ}	Z_{VBF}^{HMDJ}	c_{ggF}^{HMDJ}
Nominal	5.31	0.316	0.312
5 Vars	5.36	0.357	0.243
6 Vars	5.35	0.407	0.268
7 Vars	5.35	0.416	0.273

Table 7.5: Prediction of N_{VBF}^{HMDJ}, Z_{VBF}^{HMDJ} and c_{ggF}^{HMDJ} yields from the training sample for the three possible working points on separate BDT classifier that are predicted to yield the same N_{VBF}^{HMDJ} as the nominal cut-based categorisation.

As in the previous section each working point was used to obtain $m_{\gamma\gamma}$ distributions for the HMDJ and GGFE categories. Background estimates in the signal region were obtained using the same methodology as before. A 1^{st} order Bernstein polynomial was shown to be sufficient to model the background in all HMDJ categories and the 3^{rd} order Bernstein polynomial for all GGFE categories. The backgrounds and efficiency predictions are shown in Tables 7.6 and 7.7.

WP	$\varepsilon_{ggF}^{SR,HMDJ}$	$\delta\varepsilon_{ggF}^{SR,HMDJ}$	$\varepsilon_{VBF}^{SR,HMDJ}$	$\delta\varepsilon_{VBF}^{SR,HMDJ}$	$N_{bkg}^{SR,HMDJ}$	$\delta N_{bkg}^{SR,HMDJ}$
Nominal	0.00407	0.00005	0.1107	0.0004	59.54	3.73
5 Vars	0.00290	0.00004	0.1109	0.0006	50.34	3.44
6 Vars	0.00332	0.00004	0.1106	0.0006	39.36	3.06
7 Vars	0.00510	0.00006	0.1355	0.0006	91.74	4.08

Table 7.6: Expected signal efficiencies and background in the signal region of the HMDJ category, when using a categorisation defined by three working points on alternative BDT classifiers.

WP	$\varepsilon_{ggF}^{SR,GGFE}$	$\delta\varepsilon_{ggF}^{SR,GGFE}$	$\varepsilon_{VBF}^{SR,GGFE}$	$\delta\varepsilon_{VBF}^{SR,GGFE}$	$N_{bkg}^{SR,GGFE}$	$\delta N_{bkg}^{SR,GGFE}$
Nominal	0.3732	0.0004	0.2751	0.0006	13104.8	99.3
5 Vars	0.3744	0.0004	0.2748	0.0008	13111.2	99.3
6 Vars	0.3740	0.0004	0.2751	0.0008	13121.5	99.3
7 Vars	0.3722	0.0004	0.2502	0.0008	13066.5	99.2

Table 7.7: Expected signal efficiencies and background in the signal region of the GGFE category, when using a categorisation defined by three working points on alternative BDT classifiers.

The distributions of $\mathfrak{R}$ obtained with the pseudoexperiments, for each classifier are shown in Figure 7.6.

7.4 Discussion

The 7 variable BDT is chosen because it reduces the statistical uncertainty on the measurement of $\mathfrak{R}$, shown by narrow distribution in Figure 7.6. The working point on the 7 variable BDT provides the highest VBF signal significance. 24.0% improvement with respect to the nominal cut-based analysis and the gluon-gluon fusion contamination is reduced by 12.0%.

7.5 Results

Using the 7 variable BDT a HMDJ and GGFE is obtained. A 1^{st} order Bernstein polynomial is fitted to the data sidebands in the HMDJ category and a 3^{rd} order Bernstein polynomial is fitted

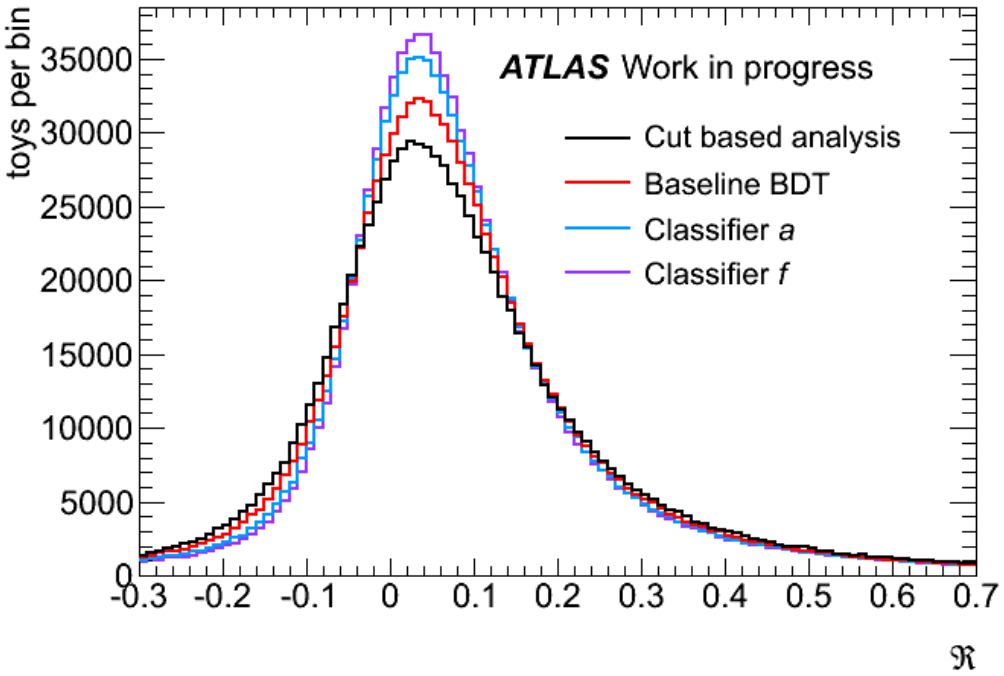


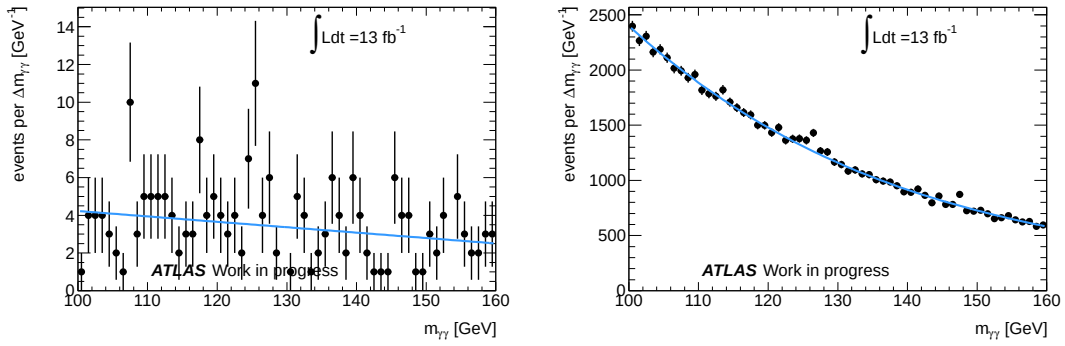
Figure 7.6: The value of $\mathfrak{R}$ predicted by random number generation for 4 alternative BDT classifiers of alternative number of variables. The colours of each histogram correspond to the colours of the crosses on the working point in Figure 7.5.

to the sidebands in the GGFE category. 35.1 background events were estimated in the HMDJ category and 13124.5 events were estimated in the GGFE category. The total number of events in the signal region of both categories are: 43 events in the HMDJ category and 13516 in the GGFE category (see Figure 7.7 and Table 7.8).

Category	N_{bkg}^{SR}	N^{SR}	N_s^{SR}
HMDJ	35.1	43	7.9
GGFE	13124.5	13516	391.5

Table 7.8: Shown for each category: the estimated background in the signal region (N_{bkg}^{SR}), the number of data events in the signal region (N^{SR}) and the estimated signal in the signal region (N_s^{SR}).

Using Equation 6.5 and the signal efficiencies obtained from the MC, the value of $\mathfrak{R}$ is measured. The statistical uncertainty associated with $\mathfrak{R}$ was determined through error propagation,



(a) HMDJ (b) GGFE

Figure 7.7: Fits to the data sidebands only, with the data points in the signal region revealed. HMDJ data events selected with the BDT chosen at the end of chapter 7. (a) HMDJ sidebands fitted with a 1^{st} order Bernstein polynomial. (b) GGFE category of data events sidebands fitted with a 3^{rd} order Bernstein polynomial.

assuming that the variables that $\mathfrak{R}$ depends on are un-correlated with one another. $\delta\mathfrak{R}$ is given by

$$\begin{aligned}\delta\mathfrak{R}^2 = {} & \delta(\varepsilon_{ggF}^{SR,GGFE})^2 \left|\frac{\partial\mathfrak{R}}{\partial(\varepsilon_{ggF}^{SR,GGFE})}\right|^2 + \delta(\varepsilon_{ggF}^{SR,HMDJ})^2 \left|\frac{\partial\mathfrak{R}}{\partial(\varepsilon_{ggF}^{SR,HMDJ})}\right|^2 + \delta(\varepsilon_{VBF}^{SR,GGFE})^2 \left|\frac{\partial\mathfrak{R}}{\partial(\varepsilon_{VBF}^{SR,GGFE})}\right|^2 \\ & + \delta(\varepsilon_{VBF}^{SR,HMDJ})^2 \left|\frac{\partial\mathfrak{R}}{\partial(\varepsilon_{VBF}^{SR,HMDJ})}\right|^2 + \delta(N_s^{SR,GGFE})^2 \left|\frac{\partial\mathfrak{R}}{\partial(N_s^{SR,GGFE})}\right|^2 + \delta(N_s^{SR,HMDJ})^2 \left|\frac{\partial\mathfrak{R}}{\partial(N_s^{SR,HMDJ})}\right|^2\end{aligned} \tag{7.1}$$

The measurement of $\mathfrak{R}$ is

$$\mathfrak{R} = 0.037 \pm 0.067 \tag{7.2}$$

This is consistent with the SM prediction of 0.075 within the statistical uncertainty.

Chapter 8

Systematic Uncertainty on Event Selection with Jets

The chosen $\mathfrak{R}$ result is dependent on two main systematic uncertainties: the selection of signal and background events in the HMDJ and GGFE categories, and the estimate of the amount of signal in each of those categories. The main systematic effects are considered in the following sections. For each systematic effect considered $\mathfrak{R}$ is recalculated using alternative methods of measure. The uncertainty for each contributions $\mathfrak{R}^{\text{effect}}_{syst}$ is defined here as half the difference between the highest value of $\mathfrak{R}$ measured for a given effect ($\mathfrak{R}^{\text{effect}}_{Max}$) and the lowest value of $\mathfrak{R}$ measure for a given effect ($\mathfrak{R}^{\text{effect}}_{Min}$)

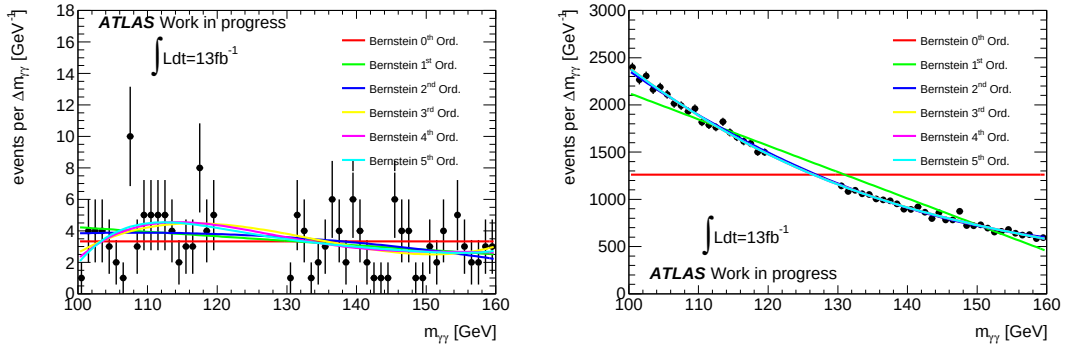
$$2 \times \delta\mathfrak{R}^{\text{effect}}_{syst} = |\mathfrak{R}^{\text{effect}}_{Max} - \mathfrak{R}^{\text{effect}}_{Min}| \tag{8.1}$$

8.1 Background Modelling

The background estimate is dependent on the fit function to the data sidebands, therefore the choice of function potentially can affect the result of $\mathfrak{R}$. In this section alternative orders of the Bernstein polynomial and other functions will be considered to fit the data sidebands of the HMDJ and LMDJ categories.

8.1.1 Different orders of Bernstein Polynomial

The criteria to choose the best Bernstein polynomial for the background estimate were described in Chapter 5. Fits to the data points are shown in Figure 8.1. It was decided upon to use a 3^{rd} order Bernstein polynomial to model the background for the GGFE category and a 1^{st} order Bernstein polynomial for the HMDJ category. Higher and lower order Bernstein polynomials would have fitted the data just as well given as the values of $q_{\hat{\nu}}$ and $p(q_{\hat{\nu}})$ were still acceptable. These values are shown in Tables 8.1 and 8.2.



(a) HMDJ (b) GGFE

Figure 8.1: Various orders of Bernstein polynomial function fitted to the data sidebands for (a) the HMDJ category, and (b) the GGFE category.

Ord.	$q_{\tilde{\nu}}$	$p(q_{\hat{\nu}})$	N_{bkg}^{SR}
0^{th}	50.5	0.375	33.4
1^{st}	47.6	0.447	35.1
2^{nd}	47.4	0.416	37.0

Table 8.1: The quality of fits (quantified by $q_{\tilde{\nu}}$ and $p(q_{\hat{\nu}})$) for 0^{th}, 1^{st} and 2^{nd} Bernstein polynomials as fit functions to the data sidebands of the HMDJ category.

Ord.	$q_{\tilde{\nu}}$	$p(q_{\hat{\nu}})$	N_{bkg}^{SR}
2^{nd}	46.2	0.507	13268.4
3^{rd}	41.4	0.667	13124.5
4^{th}	41.3	0.630	13114.0

Table 8.2: The quality of fits (quantified by $q_{\tilde{\nu}}$ and $p(q_{\hat{\nu}})$) are shown for 2^{nd}, 3^{rd} and 4^{th} Bernstein polynomials as fit functions to the data sidebands of the GGFE category.

Although the fits are still adequate, it is shown in Tables 8.1 and 8.2 that different orders in both categories give different background estimates and will therefore affect the calculation of $\mathfrak{R}$.

Three values of $\mathfrak{R}$ have been calculated with different background estimates using:

- a 0^{th} order Bernstein polynomial to fit the data sidebands of the HMDJ data points;
- a 2^{nd} order Bernstein polynomial to fit the data sidebands of the HMDJ data points;
- a 2^{nd} order Bernstein polynomial to fit the data sidebands of the GGFE data points;
- a 4^{th} order Bernstein polynomial to fit the data sidebands of the GGFE data points.

The effects are shown in Table 8.3. The systematic uncertainty due to the choice of Bernstein polynomial order is

$$\delta\mathfrak{R}^{Ord.}_{syst} = 0.028 \tag{8.2}$$

See also Figure 8.2.

Ord.	$\mathfrak{R}_i$	$\delta\mathfrak{R}_{stat}$	$\lvert\mathfrak{R}_i - \mathfrak{R}\rvert$
Chosen	0.037	0.067	-
0^{th} HMDJ	0.046	0.067	0.009
2^{nd} HMDJ	0.021	0.078	0.025
2^{nd} GGFE	0.076	0.110	0.039
4^{th} GGFE	0.036	0.065	0.001

Table 8.3: $\mathfrak{R}$ is shown for when different orders of Bernstein polynomials are fitted to the HMDJ or the GGFE category. The statistical uncertainty and the deviation from the central value ($\mathfrak{R}$) is shown for each alternative measurement.

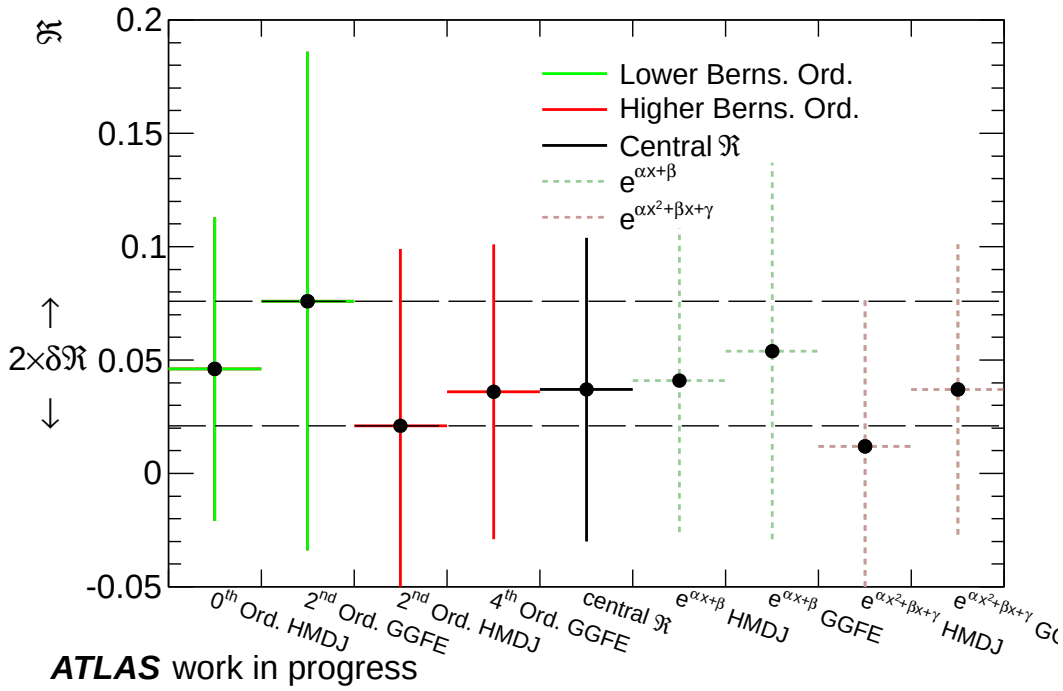


Figure 8.2: Measurement of $\mathfrak{R}$ for alternative choices of Bernstein polynomial order. Error bars show the statistical uncertainty for each measurement. $\delta\mathfrak{R}^{Ord.}_{syst}$. The extracted systematic uncertainty is shown by the dashed horizontal lines between the highest and lowest measurement of $\mathfrak{R}$.

8.1.2 Different Types of Function

As well as Bernstein polynomials, it was also shown for the ATLAS H→ γγ analysis that various exponential polynomials $e^{Pol(x)}$ can be used to model the background. A standard exponential function and exponential of a 2^{nd} order polynomial were both fitted to the sidebands of the HMDJ data points and the GGFE data points [47]. The respective values of $q_{\hat{\nu}}$ and $p(q_{\hat{\nu}})$ are shown in Tables 8.4 and 8.5, where it is shown that the values of $q_{\hat{\nu}}$ in the GGFE category are similar to

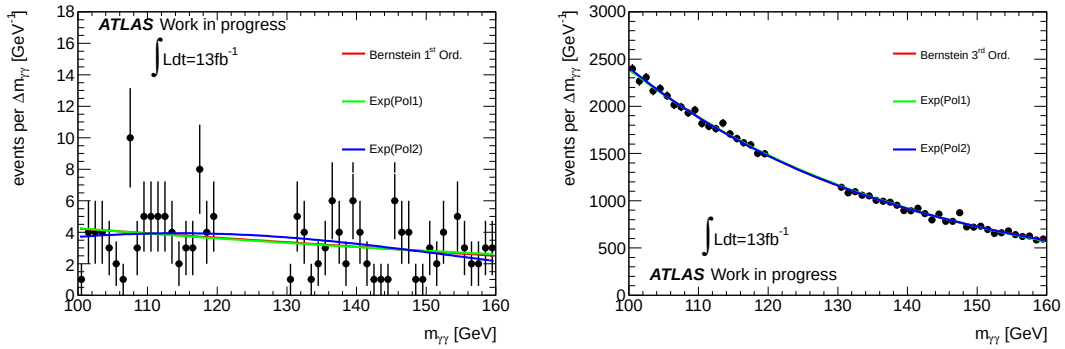
Function	$q_{\vec{\nu}}$	$p(q_{\hat{\nu}})$	N_{bkg}^{SR}
1^{st} Ord. BP	47.617	0.447	35.121
$e^{\alpha x+\beta}$	47.703	0.444	34.681
$e^{\alpha x^2+\beta x+\gamma}$	47.204	0.423	38.062

Table 8.4: The quality of fits (quantified by $q_{\vec{\nu}}$ and $p(q_{\hat{\nu}})$) for alternative functions fitted to the data sidebands of the HMDJ category.

Function	$q_{\vec{\nu}}$	$p(q_{\hat{\nu}})$	N_{bkg}^{SR}
3^{rd} Ord. BP	41.365	0.667	13124.491
$e^{\alpha x+\beta}$	43.110	0.673	13200.877
$e^{\alpha x^2+\beta x+\gamma}$	41.833	0.686	13120.967

Table 8.5: The quality of fits (quantified by $q_{\vec{\nu}}$ and $p(q_{\hat{\nu}})$) for alternative functions fitted to the data sidebands of the GGFE category.

that of the 3^{rd} order Bernstein polynomial that was fitted to the data in the GGFE category and the values of $q_{\hat{\nu}}$ in the HMDJ category are similar to that of the 1^{st} order Bernstein polynomial that was fitted to the data in the HMDJ category. It is shown in Figure 8.3 that the shapes of the exponential functions compared with Bernstein polynomials are slightly different in the HMDJ category and may therefore have a systematic impact on the background estimate.



(a) HMDJ (b) GGFE

Figure 8.3: Various functions fitted to the data sidebands for (a) the HMDJ category, and (b) the GGFE category.

The fit function should be chosen on the basis of the quality of the fit, so as to not cause any bias to the overall result. Given that the quality of exponential polynomial fits are comparable to that of the Bernstein polynomials, there is no reason for this analysis to prefer the exponential function over the Bernstein function. As a cross check, four values of $\mathfrak{R}$ have been calculated with

Function	$\mathfrak{R}_i$	$\delta\mathfrak{R}_{stat}$	$\lvert\mathfrak{R}_i - \mathfrak{R}\rvert$
Chosen	0.037	0.067	-
$e^{\alpha x+\beta}$ HMDJ	0.041	0.067	0.004
$e^{\alpha x^2+\beta x+\gamma}$ HMDJ	0.012	0.065	0.025
$e^{\alpha x+\beta}$ GGFE	0.054	0.083	0.017
$e^{\alpha x^2+\beta x+\gamma}$ GGFE	0.037	0.064	0.000

Table 8.6: $\mathfrak{R}$ for when different functions are fitted to the data sidebands in the HMDJ or the GGFE category. The statistical uncertainty and the deviation from the chosen result ($\mathfrak{R}$) is shown for each alternative measurement.

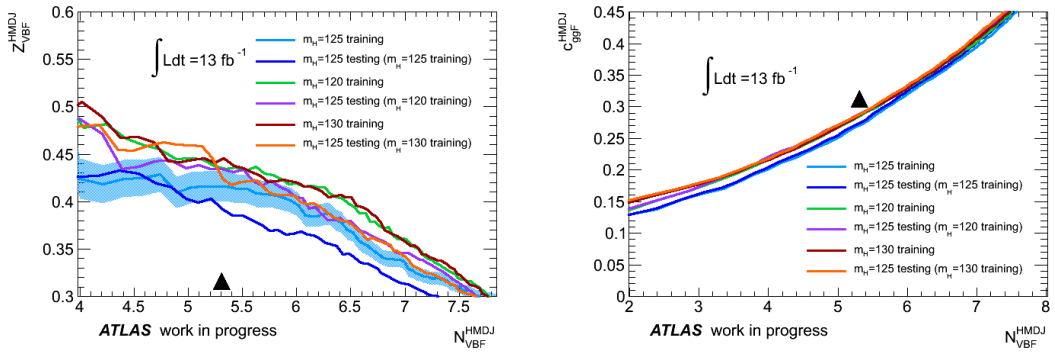
different background estimates using:

- a $e^{\alpha x+\beta}$ function to fit the data sidebands of the HMDJ category;
- a $e^{\alpha x^2+\beta x+\gamma}$ function to fit the data sidebands of the HMDJ category;
- a $e^{\alpha x+\beta}$ function to fit the data sidebands of the GGFE data category;
- a $e^{\alpha x^2+\beta x+\gamma}$ function to fit the data sidebands of the GGFE category.

The effects are shown in Table 8.6. Since it was decided not to use these functions and $\delta\mathfrak{R}^{Func.}_{syst}$ is comparable to that of $\delta\mathfrak{R}^{Ord.}_{syst}$ (see Figure 8.2) it was decided not to include $\delta\mathfrak{R}^{Func.}_{syst}$ in the total contribution.

8.2 Various m_H Signal Samples

The MC signal samples that were used to train the MVA classifier were generated with a Higgs boson mass of $m_H = 125\,\text{GeV}$. However, the mass of the Higgs boson being analysed hasn't yet been determined exactly, so training with these samples could have biased the result. To ensure that there is no bias, the classifier has been re-trained with a MC signal sample of $m_H = 120\,\text{GeV}$, and again re-trained with a MC signal sample of $m_H = 130\,\text{GeV}$. Everything else in the training was kept in the exact same way as previously described. The two classifiers above were then tested using the same signal and background testing samples that were used to test the chosen BDT, with $m_H = 125\,\text{GeV}$. The VBF signal yields, gluon-gluon fusion contamination and significance in the HMDJ (N^{HMDJ}_{VBF}, Z^{HMDJ}_{VBF} and c^{HMDJ}_{ggF}) were then calculated for a variety of working points and the Z^{HMDJ}_{VBF} vs N^{HMDJ}_{VBF} and the c^{HMDJ}_{ggF} vs N^{HMDJ}_{VBF} curves are shown in Figure 8.4. For reference, the



(a) Z_{VBF}^{HMDJ} vs N_{VBF}^{HMDJ} (b) c_{ggF}^{HMDJ} vs N_{VBF}^{HMDJ}

Figure 8.4: Performance of BDT on the testing sample, which was trained using VBF signal samples of $m_H = 120\,\text{GeV}$ and $m_H = 130\,\text{GeV}$. (a) Z_{VBF}^{HMDJ} vs c_{ggF}^{HMDJ} and (b) c_{ggF}^{HMDJ} vs N_{VBF}^{HMDJ}.

Training Mass	N_{VBF}^{HMDJ}	Z_{VBF}^{HMDJ}	c_{ggF}^{HMDJ}
$m_H = 120\,\text{GeV}$	5.31869	0.423838	0.288388
$m_H = 125\,\text{GeV}$	5.31105	0.434425	0.288282
$m_H = 130\,\text{GeV}$	5.35109	0.416128	0.272978

Table 8.7: N_{VBF}^{HMDJ}, Z_{VBF}^{HMDJ} and c_{ggF}^{HMDJ} for working points where the BDT has been trained separately for signal samples generated with a Higgs mass of 120 GeV, 125 GeV and 130 GeV.

performance of the chosen MVA is also shown in Figure 8.4. The working points that were chosen (WP_{120} and WP_{130}), were the ones that yielded a similar performance to that of the working point from the $m_H = 125\,\text{GeV}$ training (WP_{125}). The yields that were obtained from the training sample are shown in Table 8.7. Using the selection criteria of WP_{120} and WP_{130} one obtains alternative HMDJ and GGFE categories that were fitted with functions to estimate the background in the signal region. It was evaluated that a 1^{st} order Bernstein polynomial best fitted the sideband data points in the HMDJ category and a 3^{rd} order Bernstein polynomial best fitted the sideband data points in the GGFE category. $\mathfrak{R}$ has been re-calculated for these working points and the relative variation on $\mathfrak{R}$ is shown in Table 8.8 and Figure 8.5.

It is shown in Figure 8.5 that the central measured value of $\mathfrak{R}$ ($m_H = 125\,\text{GeV}$) is the lowest measured value, however the statistical error bars are fairly large. The variation in the $\mathfrak{R}$ measurements is therefore likely to be of statistical origin. Irrespective of this, the associated systematic $\delta\mathfrak{R}_{syst}^{m_H}$ is conservatively estimated to be half the difference between the highest value of $\mathfrak{R}$ ($m_H = 120\,\text{GeV}$) and the lowest value of $\mathfrak{R}(m_H = 125\,\text{GeV})$.

$$\delta\mathfrak{R}^{m_H}_{syst} = 0.019 \tag{8.3}$$

Working Point	$\mathfrak{R}_i$	$\delta\mathfrak{R}_{stat}$	$\lvert\mathfrak{R}_i - \mathfrak{R}\rvert$
Chosen	0.037	0.067	-
WP $_{120}$ GeV	0.074	0.074	0.037
WP $_{130}$ GeV	0.056	0.071	0.019

Table 8.8: Individual $\mathfrak{R}$ values measured ($\mathfrak{R}_i$) using BDTs trained separately with signal samples with a Higgs mass of 120 GeV, 125 GeV or 130 GeV. The statistical uncertainty ($\delta\mathfrak{R}_{stat}$) and the deviation from the nominal $\mathfrak{R}$ value are also shown for each alternative measurement.

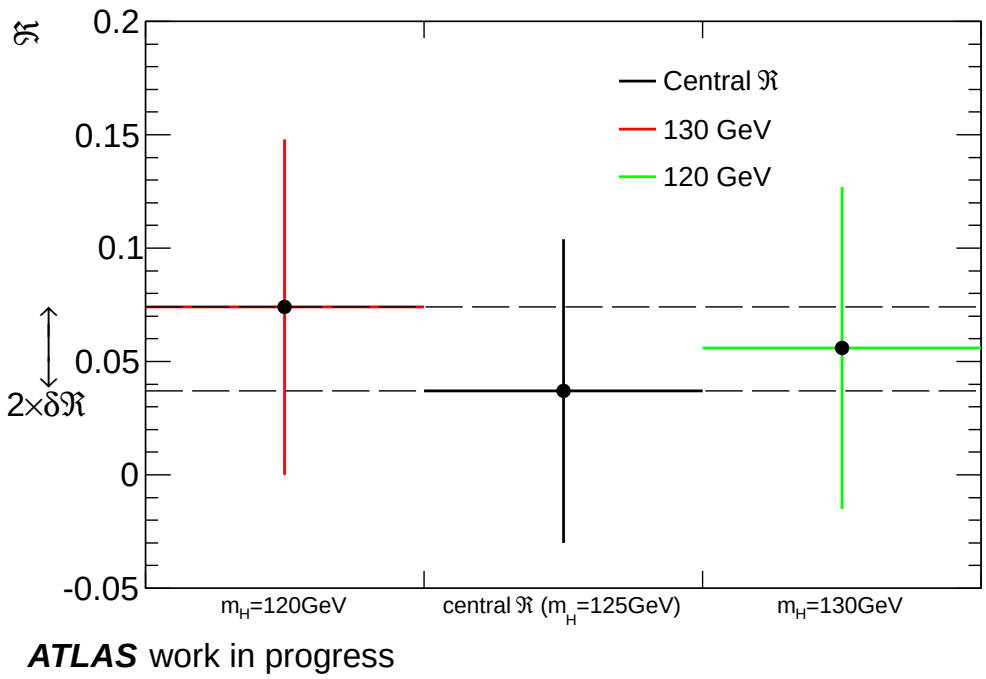


Figure 8.5: Measurements of $\mathfrak{R}$ obtained with BDTs trained separately for signal samples with a Higgs mass of 120 GeV, 125 GeV (nominal) and 130 GeV. Error bars show the statistical uncertainty for each measurement. $\delta\mathfrak{R}^{m_H}_{syst}$ is indicated by the horizontal dashed lines between the highest and lowest measurements of $\mathfrak{R}$.

8.3 Jet Energy Scale (JES) and Jet Energy Resolution (JER) Uncertainties

The energy of the jets is calibrated using the EM+JES scheme and the jet energy resolution is also corrected to agree with MC. Both of these effects have associated uncertainties (refer to Section 4). Since measurements of the jet energy play an important role in event classification ($p_{T,j1}$, $p_{T,j2}$, and M_{jj} are used for the HMDJ classification), knowing the uncertainty affects how many signal events can potentially get selected into one category or another.

8.3.1 JES Uncertainty

The uncertainty is calculated $in-situ$ taking into account the correlations of the systematic parameters. The uncertainty due to each JES is determined using the follow recipe. The event categorisation is run as normal and will yield of N^c signal events in each category. The categorisation was then repeated but this time after the E_{jet} and $p_{T,jet}$ of the jets were scaled up by and uncertainty factor u, which is dependent on a variety of systematic parameters

$$E_{jet}^{new} = E_{jet}^{old}(1+u) \tag{8.4}$$

$$p_{T,jet}^{new} = p_{T,jet}^{old}(1+u) \tag{8.5}$$

This will be referred to as JESup. This will now yield a different number of signal events for each category N^c_{JESup}. The categorisation was repeated again but this time after the E_{jet} and $p_{T,jet}$ of each jet were scaled down

$$E_{jet}^{new} = E_{jet}^{old}(1-u) \tag{8.6}$$

$$p_{T,jet}^{new} = p_{T,jet}^{old}(1-u) \tag{8.7}$$

yielding N^c_{JESup} signal events for each category.

The response of the jets as discussed in Section 4.2 can be affected by a variety of factors:

- **Baseline** overall measurement of the JES uncertainty;
- **High** $|\eta|$ different amount of material and technology means that the jet response can vary

depending on its direction. Since forward jets are used in this analysis the uncertainty is largely dependent on this;

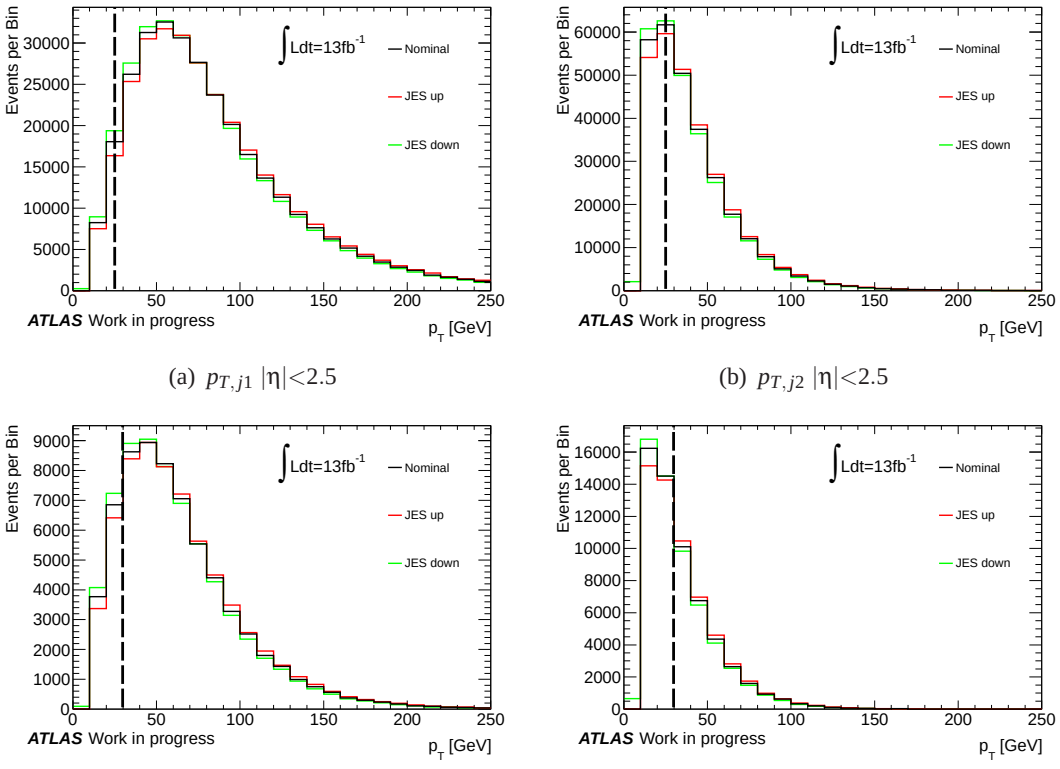
- **Flavour** the jet response can vary depending on whether the fragmentation was initiated by quarks or gluons;
- μ the jet response can vary depending on mean number of interactions per crossing;
- N_{PV} the jet response can vary with respect to the number of primary vertices in the bunch crossing;
- *b*-**Jet** the jet response can vary depending on whether the fragmentation was initiated by a *b*-quark;
- **Close-by** the jet response can vary depending on whether the jet is close (in ΔR) to another jet.

By way of example, the effects of u on $p_{T,j1}$ and $p_{T,j2}$ from the forward component ("high $|\eta|$") of the calculation are shown in Figures 8.6(a) and 8.6(b) for the barrel region $|\eta| < 2.5$ and Figures 8.6(c) and 8.6(d) for the jets in the end caps. These plots show p_T binned for every event in the VBF signal MC which contain at least two tag jets where the p_T thresholds have been lowered to 15 GeV. The black lines show the p_T thresholds for normal tag jet classification. In the JES downward scaling, the green distribution is skewed to lower energies indicating that more events would be less likely to have two tag jets and therefore HMDJ events would be more likely to be selected into the GGFE category. In the JES upward scaling, the distribution is skewed to higher energies and the opposite effect occurs.

The systematic uncertainty, α due to upwards and downwards scaling is given by the difference in efficiencies with respect to the unscaled selection efficiency:

$$\alpha_{JESup(down)} = \frac{N^c_{JESup(down)} - N^c}{N^c} \tag{8.8}$$

α_{JESup} is shown in Table 8.9, $\alpha_{JESdown}$ is shown in Table 8.10.



(a) $p_{T,j1}$ $|\eta|<2.5$

(b) $p_{T,j2}$ $|\eta|<2.5$

(c) $p_{T,j1}$ $|\eta|>2.5$

(d) $p_{T,j2}$ $|\eta|>2.5$

Figure 8.6: Effect of the jet energy scale systematic uncertainty on the distributions of the p_T of the tag jets in the signal. The distributions are shown without correction ("nominal") as well as with the jet energy scalings described in the text. (a) Leading jet in the barrel, (b) Subleading jet in the barrel, (c) Leading jet in the end-caps and (d) Subleading jet in the end-caps.

JESup	HMDJ		GGFE	
	ggF	VBF	ggF	VBF
Baseline	0.097	0.068	-0.001	-0.019
High $\|\eta\|$	0.101	0.077	-0.001	-0.014
Flavour	0.093	0.057	-0.000	-0.005
μ	0.018	0.020	-0.000	-0.003
N_{PV}	0.014	0.012	-0.000	-0.000
b-jet	0.000	0.000	-0.001	-0.017
Close by jets	0.010	0.008	-0.000	-0.002

Table 8.9: α shown for various jet energy scale contributions when the energies on the jets is scaled up for the VBF and gluon-gluon fusion production mechanisms, in both HMDJ and GGFE categories.

JESdown	HMDJ		GGFE	
	ggF	VBF	ggF	VBF
Baseline	-0.079	-0.042	0.001	0.013
High $\|\eta\|$	-0.089	-0.052	0.001	0.012
Flavour	-0.081	-0.042	0.000	0.001
μ	-0.011	-0.001	0.000	0.001
N_{PV}	-0.012	-0.004	0.000	0.000
b-jet	-0.000	-0.000	0.001	0.011
Close by jets	-0.012	-0.006	0.000	0.002

Table 8.10: α shown for various jet energy scale contributions when the energies on the jets is scaled down for the VBF and gluon-gluon fusion production mechanisms, in both HMDJ and GGFE categories.

Using the scaled efficiencies $\mathfrak{R}$ has been recalculated. The resulting values of $\mathfrak{R}$ from α_{JESup} are shown in Table 8.11 and the resulting values of $\mathfrak{R}$ from $\alpha_{JESdown}$ are shown in Table 8.12 (see also Figure 8.7). The systematic uncertainty due to the jet energy scale is:

$$\delta\mathfrak{R}^{JES}_{syst} = 0.006 \tag{8.9}$$

see Figure 8.7.

JESup	$\mathfrak{R}_i$	$\delta\mathfrak{R}_{stat}$	$\lvert\mathfrak{R}_i - \mathfrak{R}\rvert$
Chosen	0.037	0.067	-
Baseline	0.032	0.063	0.005
High $\lvert\eta\rvert$	0.032	0.063	0.005
Flavour	0.033	0.064	0.004
μ	0.036	0.066	0.001
N_{PV}	0.037	0.066	0.000
b-jet	0.038	0.067	0.001
Close by jets	0.037	0.067	0.005

Table 8.11: $\mathfrak{R}$, statistical uncertainty and relative systematic error on $\mathfrak{R}$ for various jet energy scale contributions when the energy of the jets is scaled up.

JESdown	$\mathfrak{R}_i$	$\delta\mathfrak{R}_{stat}$	$\lvert\mathfrak{R}_i - \mathfrak{R}\rvert$
Chosen	0.037	0.067	-
Baseline	0.041	0.070	0.004
High $\lvert\eta\rvert$	0.043	0.071	0.006
Flavour	0.042	0.070	0.005
μ	0.038	0.067	0.001
N_{PV}	0.038	0.068	0.001
b-jet	0.038	0.067	0.001
Close by jets	0.038	0.068	0.001

Table 8.12: $\mathfrak{R}$, statistical uncertainty and relative systematic error on $\mathfrak{R}$ for various jet energy scale contributions when the energy of the jets is scaled down.

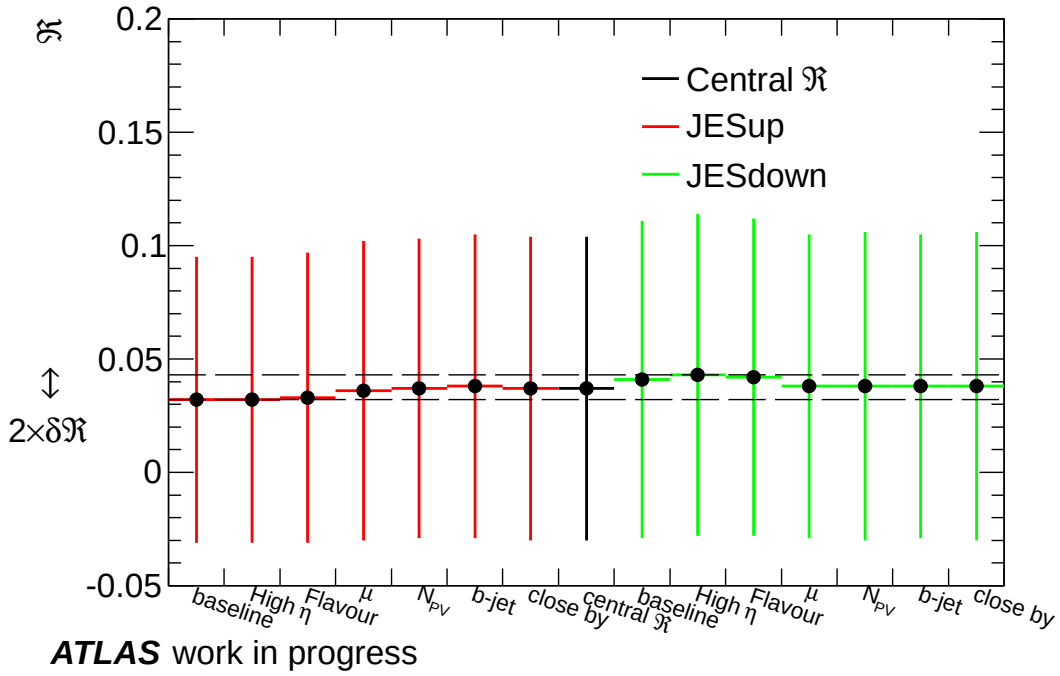


Figure 8.7: Measurement of $\mathfrak{R}$ for JESup and JESdown. Error bars show the statistical uncertainty for each measurement. $\delta\mathfrak{R}^{JES}_{syst}$ is shown by the dashed horizontal lines between the highest and the lowest values obtained.

8.3.2 JER Uncertainty

The jet energy resolution uncertainty is calculated using in-situ techniques, the total systematic uncertainty on the event selection is calculated in a similar way to the jet energy scale. The p_T and E of the jets are smeared by a factor u, which is obtained from a Gaussian random number generator for each event, using the JER uncertainty as 1σ.

$$E^{new}_{jet} = E^{old}_{jet}(1+u) \tag{8.10}$$

$$p^{new}_{T,jet} = p^{old}_{T,jet}(1+u) \tag{8.11}$$

The effects of JER uncertainties on the HMDJ category and the GGFE category for the gluon-gluon fusion and VBF signals are summarised in Table 8.13 . $\delta\mathfrak{R}^{JER.}_{syst}$ is calculated as 0.001.

Systematic	ggF	VBF
HMDJ	-0.000	0.033
GGFE	0.028	0.020

Table 8.13: α is calculated due to the jet energy resolution for the VBF and gluon-gluon fusion production mechanisms, in both HMDJ and GGFE categories.

8.4 Other Signal Contributions

Another systematic uncertainty may arise from the VH and ttH signal contributions, that are not completely negligible. Assuming SM predictions, the VH and ttH signal contributions will account for 3.6% and 0.6% of the signal in the GGFE and HMDJ categoies, respectively. The value of $\mathfrak{R}$ was then re-calculated under the assumption that these additional contributions were additional backgrounds and could be subtracted off. $\delta\mathfrak{R}^{VH,ttH}_{syst}$ was calculated as 0.003.

8.5 Uncertainty due to Limited Data

In Section 6.4, toy MC experiments were used to investigate the measurement of $\mathfrak{R}$ with 13 fb^{-1} of data. 1,000,000 toy experiments were generated for each one of five different scenarios with alternative cross sections for gluon-gluon fusion and VBF. The distributions of the resulting $\mathfrak{R}$ values are shown in Figure 6.5. It can be seen that the distributions are not necessarily symmetric, and that they centre approximately (but not exactly) on the true value of $\mathfrak{R}$. A potential systematic uncertainty is therefore associated with this effect. In the same section it was verified that, as expected, increasing the size of the data set will reduce this systematic effect. With larger data sets, the distributions become more symmetric and their centre will converge on to the true value of $\mathfrak{R}$. Nevertheless, as the present measurement is obtained from a limited data set of 13 fb^{-1} the corresponding systematic uncertainty has to be quantified.

In order to estimate the systematic uncertainty, three scenarios were investigated:

1. Assume the SM gluon-gluon fusion and VBF cross sections and SM $H \to \gamma\gamma$ branching ratio;

2. Same as 1 except $\sigma_{ggF} \to \sigma_{ggF} \times 1.6$;

3. Same as 1 except $\sigma_{VBF} \to \sigma_{VBF} \times 1.7$.

Scenario	$\mathfrak{R}_{true}$	$\mathfrak{R}_{peak}$	$\mathfrak{R}_{med}$	$\mathfrak{R}_{peak}-\mathfrak{R}_{true}$	$\mathfrak{R}_{med}-\mathfrak{R}_{true}$
1	0.075	0.037	0.060	-0.038	-0.015
2	0.048	0.031	0.046	-0.017	-0.002
3	0.121	0.070	0.104	-0.051	-0.016

Table 8.14: Peak and median values associated with the distributions of 1,000,000 MC toy experiments for 3 alternative scenarios of cross sections. The difference between these values and the true value of $\mathfrak{R}$ is also shown.

Scenario 1 has been chosen as recent ATLAS measurements [24, 28, 26] show that the observable properties of the Higgs boson are consistent with the Standard Model prediction. However, although the measurements are consistent with the Standard Model within the experimental uncertainties, Scenarios 2 and 3 were also investigated, based on the signal strength measurements (μ) of various production mechanisms. These results were as follows

$$\begin{aligned} \mu_{ggF+ttH} \times \frac{B}{B_{SM}} &= 1.6 \pm^{0.3}_{0.3} (\text{stat}) \pm^{0.3}_{0.2} (\text{syst}) \\ \mu_{VBF} \times \frac{B}{B_{SM}} &= 1.7 \pm^{0.8}_{0.8} (\text{stat}) \pm^{0.5}_{0.4} (\text{syst}) \end{aligned} \tag{8.12}$$

[24], where B is the diphoton branching fraction.

For each of these scenarios 1,000,000 toy experiments were run. The optimised MVA working point that was determined in Section 7.3 was used for all scenarios. The distributions of the toy experiments for these scenarios are shown in Figure 8.8 with the centre (median) positions and peak positions for each distribution[1]. An asymmetry is observed in each distribution, and as a consequence the median and peak positions are not equivalent. The comparison between the true $\mathfrak{R}$ ($\mathfrak{R}_{true}$) value in each scenario, with the median ($\mathfrak{R}_{med}$) and peak position ($\mathfrak{R}_{peak}$) is shown in Table 8.14.

Due to the asymmetric nature of the distributions, the estimated systematic uncertainty is taken to be half the difference between the median and the true value of $\mathfrak{R}$. To be conservative the maximum difference is chosen, which occurs in Scenario 3. The uncertainty due to this effect is therefore

$$\delta\mathfrak{R}^{\mathcal{L}}_{syst} = 0.008 \tag{8.13}$$

[1] The peak position was estimated by fitting a Gaussian function to five bins either side of the central bin.

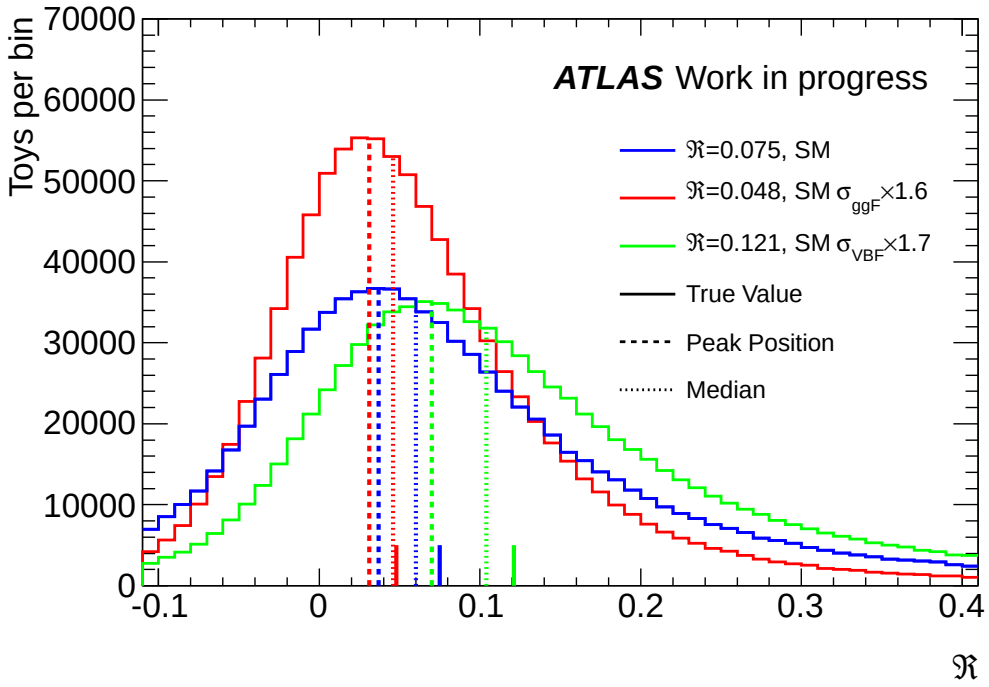


Figure 8.8: Distributions of $\mathfrak{R}$ with 1,000,000 toys of pseudodata of 3 different cross section scenarios. The Value of $\mathfrak{R}$ in each toy is calculated from randomly generated numbers that are consistent with the expectation of signal and background for each cross section scenario and using the chosen MVA working point.

8.6 Total Systematic Uncertainty

Out of all the contributions considered, the largest effect was due to the choice of Bernstein polynomial order. Using all of the contributions the total systematic effect is calculated by adding together in quadrature:

$$\begin{aligned} \delta\mathfrak{R}_{syst} &= \delta\mathfrak{R}^{Ord.}_{syst} \oplus\delta\mathfrak{R}^{m_H}_{syst} \oplus\delta\mathfrak{R}^{JES}_{syst} \oplus\delta\mathfrak{R}^{JER}_{syst} \oplus\delta\mathfrak{R}^{\mathcal{L}}_{syst} \oplus\delta\mathfrak{R}^{VH,ttH}_{syst} \\ &= 0.028 \oplus 0.019 \oplus 0.006 \oplus 0.001 \oplus 0.008 \oplus 0.003 \\ &= 0.035 \end{aligned} \tag{8.14}$$

Chapter 9

Conclusion

The Higgs boson is the last remaining particle in the Standard Model to be found. A new boson has recently been observed displaying properties that are consistent with the Standard Model Higgs boson. It is important to measure the production rates of the Higgs boson, as this will unlock information about the couplings to other particles, as explained in Chapter 1. This thesis has investigated reoptimising the event selection, to categorise the $H \rightarrow \gamma\gamma$ events that are produced by the VBF processes.

The original category developed by ATLAS to be enriched in VBF events (the HMDJ category) was a cut-based approach. The studies in Chapter 5 showed that the amount of signal expected in this category is limited, and at the same time many background events and nearly as much gluon-gluon fusion signal was contaminating this category. In this thesis a boosted decision tree (BDT) was investigated to improve on this categorisation scheme. Using five input variables (p_{Tj1}, p_{Tj2}, $\Delta\eta_{jj}$, M_{jj} and $|\vec{p}_{T,\gamma\gamma} + \vec{p}_{T,jj}|$) in the BDT, an improvement on the selection performance was achieved with respect to the cut-based approach. The VBF signal efficiency increased by 22.6%, for a significance similar to that of the cut-based approach. The significance increased by 5.9%, for a VBF signal efficiency similar to that of the cut-based approach and at the same time, the contamination from the gluon-gluon fusion signal was decreased by 21.7%.

The VBF rate relative to the VBF and gluon-gluon fusion rates ($\mathfrak{R}$) was measured to check for consistency with the Standard Model. The amount of signal from the HMDJ category and a category rich in gluon-gluon fusion $H \rightarrow \gamma\gamma$ events (the GGFE category) were required. This was obtained using a background subtraction method described in Chapter 6. It was shown a 1^{st} order

Bernstein polynomial best fit the background distribution in the HMDJ category and a 3^{rd} order polynomial polynomial best fit the background distribution in the GGFE category.

In Chapter 7, it was demonstrated with pseudodata experiments that a HMDJ category which gives a higher VBF signal significance is likely to measure $\mathfrak{R}$ with a lower statistical uncertainty. The pseudodata experiments also showed that the statistical uncertainty is large with the amount of data currently available. To improve the measurement of $\mathfrak{R}$ it was decided to increase the VBF signal significance by including additional variables $p_{T,t\gamma\gamma}$ and $p_{T,\gamma 1}$. This new BDT provided an improvement in VBF signal significance by 24.0% relative to the cut-based analysis and still selects 12.0% less gluon-gluon fusion signal than the cut-based approach.

The appropriate measurement procedures were investigated in chapter 8 to assess the systematic uncertainty. $\mathfrak{R}$ is measured as

$$\mathfrak{R} = 0.037 \pm 0.067(\text{stat}) \pm 0.035(\text{syst})$$

to be compared with a Standard Model prediction of $\mathfrak{R} = 0.075$. Various measurements have been carried out by ATLAS on this new particle, such as decay channel rates, spin measurements and mass measurements. All results have shown that this new particle is consistent with a Higgs boson as predicted by the Standard Model within the current precision of the tests. The measurement of $\mathfrak{R}$ provided by this thesis also suggests consistency with the Standard Model, within the measured uncertainty.

Bibliography

[1] ATLAS Collaboration, *Signal studies for $H \to \gamma\gamma$*, ATL-COM-PHYS-2012-501, 2012, https://cds.cern.ch/record/1447436/files/ATL-COM-PHYS-2012-501.pdf.

[2] ATLAS Collaboration, *Improved analysis of the search for the Higgs boson decaying to two photons with 4.9 fb^{-1}*, ATL-COM-PHYS-2012-502, 2012, https://cds.cern.ch/record/1447437/files/ATL-COM-PHYS-2012-502.pdf.

[3] ATLAS Collaboration, *Observation of an excess of events in the search for the Standard Model Higgs boson in the $\gamma\gamma$ channel with the ATLAS detector*, ATL-COM-PHYS-2012-788, 2012, https://cds.cern.ch/record/1454934/files/ATL-COM-PHYS-2012-788.pdf.

[4] ATLAS Collaboration, *Statistics studies for $H \to \gamma\gamma$ search for 3.24 fb^{-1} of 2012 dataset and combined results with the full 2011 dataset*, ATL-COM-PHYS-2012-757, 2012, https://cds.cern.ch/record/1453773/files/ATL-COM-PHYS-2012-757.pdf.

[5] ATLAS Collaboration, *Signal studies for $H \to \gamma\gamma$ - 8 TeV*, ATL-COM-PHYS-2012-755, 2012, https://cds.cern.ch/record/1453770/files/ATL-COM-PHYS-2012-755.pdf.

[6] ATLAS Collaboration, *Statistics studies for $H \to \gamma\gamma$ in the full 2011 dataset*, ATL-COM-PHYS-2012-732, 2012, https://cds.cern.ch/record/1453249/files/ATL-COM-PHYS-2012-732.pdf.

[7] ATLAS Collaboration, *Search for the Higgs Boson in the diphoton decay Channel with Data Collected at 7 and 8 TeV*, ATL-COM-PHYS-2012-503, 2012, https://cds.cern.ch/record/1447438/files/ATL-COM-PHYS-2012-503.pdf.

[8] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. **B 716** (1) (2012) 1 – 29.

[9] ATLAS Collaboration, *Observation of an excess of events in the search for the Standard Model Higgs boson in the $\gamma\gamma$ channel with the ATLAS detector*, ATLAS-CONF-2012-091, 2012, http://cds.cern.ch/record/1460410/files/ATLAS-CONF-2012-091.pdf.

[10] ATLAS Collaboration, *Combined measurements of the mass and signal strength of the Higgs-like boson with the ATLAS detector using up to 25 fb^{-1} of proton-proton collision data*, ATLAS-CONF-2013-014, 2013, http://cds.cern.ch/record/1523727/files/ATLAS-CONF-2013-014.pdf.

[11] D. Griffiths, *Introduction to Elementary Particles*, John Wiley & Sons, 1987.

[12] G. 't Hooft, M. Veltman, *Regularization and Renormalization of Gauge Fields*, Nucl. Phys. **B 44** (1972) 189–213.

[13] T. W. B. Kibble, *Symmetry Breaking in Non-Abelian Gauge Theories*, Phys. Rev. **155** (1967) 1554–1561.

[14] P. W. Higgs, *Spontaneous Symmetry Breakdown without Massless Bosons*, Phys. Rev. **145** (1966) 1156–1163.

[15] G. Guralnik, C. Hagen, T. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13** (1964) 585–587.

[16] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508–509.

[17] P. W. Higgs, Broken symmetries, massless particles and gauge fields, Phys. Lett. **12** (1964) 132–133.

[18] F. Englert, R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321–323.

[19] J. Beringer, et al., *Review of Particle Physics*, Phys. Rev. **D 86** (2012) 010001.

[20] E. A. Paschos, *Electroweak Theory*, Cambridge University Press, 2007.

[21] C. Mariotti, G. Passarino, R. Tanaka, et al., *Handbook of LHC Higgs Cross Sections: 2. Differential Distributions*, CERN-2012-002, (2012), [hep-ph]. arXiv:1201.3084.

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

[23] CMS Collaboration, *Combination of Standard Model Higgs boson searches and measurements of the properties of the new boson with a mass near 125 GeV*, CMS-PAS-HIG-12-045, 2012, https://cdsweb.cern.ch/record/1494149/files/HIG-12-045-pas.pdf.

[24] ATLAS Collaboration, *Measurements of the properties of the Higgs-like boson in the two photon decay channel with the ATLAS detector using 25* fb^{-1} *of proton-proton collision data*, ATLAS-CONF-2013-012, 2013, http://cds.cern.ch/record/1523698/files/ATLAS-CONF-2013-012.pdf.

[25] ATLAS Collaboration, *Measurements of the properties of the Higgs-like boson in the four lepton decay channel with the ATLAS detector using 25* fb^{-1} *of proton-proton collision data*, ATLAS-CONF-2013-013, 2013, http://cds.cern.ch/record/1523699/files/ATLAS-CONF-2013-013.pdf.

[26] ATLAS Collaboration, *Combined coupling measurements of the Higgs-like boson with the ATLAS detector using up to 25* fb^{-1} *of proton-proton collision data*, ATLAS-CONF-2013-034, 2013, http://cds.cern.ch/record/1528170/files/ATLAS-CONF-2013-034.pdf.

[27] CMS Collaboration, *Combination of standard model Higgs boson searches and measurements of the properties of the new boson with a mass near 125 GeV*, CMS-PAS-HIG-13-005, 2013.

[28] ATLAS Collaboration, *Study of the spin of the new boson with up to 25* fb^{-1} *of ATLAS data*, ATLAS-CONF-2013-040, 2013, http://cds.cern.ch/record/1542341/files/ATLAS-CONF-2013-040.pdf.

[29] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[30] ATLAS Collaboration, *ATLAS Experiment, Public Results*, (2013)https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults.

[31] M. Marcisovsky, *Operational experience of the ATLAS pixel detector*, JINST **6** (2011) C12001.

[32] A. Andreazza, *Offline calibrations and performance of the ATLAS Pixel Detector*, ATL-INDET-PROC-2011-015, 2011, http://cds.cern.ch/record/1384158/files/ATL-INDET-PROC-2011-015.pdf.

[33] R. L. Bates, et al., *The ATLAS SCT grounding and shielding concept and implementation*, JINST **7** (2012) P03005.

[34] N. Nikiforou, *Performance of the ATLAS Liquid Argon Calorimeter after three years of LHC operation and plans for a future upgrade*, 2013[hep-ex]. arXiv:1306.6756.

[35] J. P. Archambault, et al., *Performance of the ATLAS liquid argon forward calorimeter in beam tests*, JINST **8** (2013) P05006.

[36] E. Diehl, *Calibration and Performance of the ATLAS Muon Spectrometer*, ATL-MUON-PROC-2011-004, 2011, http://cds.cern.ch/record/1385884/files/ATL-MUON-PROC-2011-004.pdf.

[37] ATLAS Collaboration, *Search for the Standard Model Higgs boson in the diphoton decay channel with 4.9 fb^{-1} of ATLAS data at $\sqrt{s} = 7\,TeV$*, ATLAS-CONF-2011-161, 2011, http://cds.cern.ch/record/1406356/files/ATLAS-CONF-2011-161.pdf.

[38] T. Sjostrand, S. Mrenna, P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **0605** (2006) 026.

[39] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys, J. **C 70** (2010) 823–874.

[40] T. Sjostrand, S. Mrenna, P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852–867.

[41] S. Alioli, P. Nason, C. Oleari, E. Re, *NLO Higgs boson production via gluon fusion matched with shower in POWHEG*, JHEP **0904** (2009) 002.

[42] P. Nason, C. Oleari, *NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG*, JHEP **1002** (2010) 037.

[43] S. Agostinelli, et al., *GEANT4: A Simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250–303.

[44] ATLAS Collaboration, *Expected performance of the ATLAS experiment: Reconstruction and Identification of Photons*, ATL-PHYS-PUB-2011-007 2009, http://cds.cern.ch/record/1345329/files/ATL-PHYS-PUB-2011-007.pdf.

[45] ATLAS Collaboration, *Expected photon performance in the ATLAS experiment*, ATL-PHYS-PUB-2011-007, 2011, http://cds.cern.ch/record/1345329/files/ATL-PHYS-PUB-2011-007.pdf.

[46] ATLAS Collaboration, *Expected performance of the ATLAS experiment: detector, trigger and physics*, CERN-OPEN-2008-020, 2009, http://cds.cern.ch/record/1125884/files/CERN-OPEN-2008-020.pdf.

[47] ATLAS Collaboration, *Observation and study of the Higgs boson candidate in the two photon decay channel with the ATLAS detector at the LHC*, ATLAS-CONF-2012-168, 2012, http://cds.cern.ch/record/1499625/files/ATLAS-CONF-2012-168.pdf.

[48] ATLAS Collaboration, *Measurement of the inclusive isolated prompt photon cross section in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, Phys. Rev. **D 83** (2011) 052005.

[49] M. Cacciari, G. P. Salam, G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **0804** (2008) 063.

[50] ATLAS Collaboration, *Selection of jets produced in proton-proton collisions with the ATLAS detector using 2011 data*, ATLAS-CONF-2012-020, 2012, http://cds.cern.ch/record/1430034/files/ATLAS-CONF-2012-020.pdf.

[51] T. Barillari, *Jet Energy Scale Uncertainties in ATLAS*, J Phys.: Conf. Ser. **404** (2012) 012012.

[52] ATLAS Collaboration, *Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 7$ TeV*, Eur. Phys. J. **C 73** (2013) 2304, [hep-ex]. arXiv:1112.6426.

[53] ATLAS Collaboration, *Study of jets produced in association with a W boson in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, Phys. Rev. **D 85** (2012) 092002.

[54] ATLAS Collaboration, *Expected electron performance in the ATLAS experiment*, ATL-PHYS-PUB-2011-006, 2011, http://cds.cern.ch/record/1345327/files/ATL-PHYS-PUB-2011-006.pdf.

[55] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. **B 716** (2012) 30 – 61.

[56] ATLAS Collaboration, *Search for the Standard Model Higgs boson in the diphoton decay channel with 4.9 fb^{-1} of pp collisions at $\sqrt{s} = 7$ TeV with ATLAS*, Phys. Rev. Lett. **108** (2012) 111803.

[57] DESY, *Statistical Error of Efficiency Determination from Weighted Events*, (2009)www.desy.de/ blist/notes/effic.ps.gz.

[58] A. Hocker, J. Stelzer, F. Tegenfeldt, et al., *TMVA - Toolkit for Multivariate Data Analysis*, PoS ACAT (2007) 040. arXiv:physics/0703039.

[59] B. P. Roe, et al., *Boosted Decision Trees, an alternative to artificial neural networks*, Nucl. Instrum. Meth. **A 543** (2005) 577–584.

[60] F. James, M. Winkler, *Minuit User's Guide* 2004, http://seal.web.cern.ch/seal/documents/minuit/mnusersguide.pdf.

[61] G. Cowan, *Numerical Error Propagation*, (2011)www.pp.rhul.ac.uk/∼cowan/stat/error_prop.ps.

[62] S. Caron, G. Cowan, et al., *Absorbing systematic effects to obtain a better background model in a search for new physics*, JINST **4** (2009) P10009.

[63] G. Cowan, *Statistical Data Analysis*, Oxford University Press, 1998.