Abstract

The discovery or exclusion of the Higgs Boson is one of the main goals in today’s high energy physics experiments. In the Atlas experiment at the Large Hadron Collider at CERN in Geneva, one of the most promising discovery modes is the Higgs decay to two $W$-bosons. A search optimization has been performed using the key observables of this Higgs decay channel, applying advanced techniques such as artificial neural networks and boosted decision tree methods. Results on the signal significance with the optimized Higgs candidate selection will be presented. The potential for the discovery of the Higgs by Atlas using the $W$-decay channel is discussed.
Supervisor:
Dr. Max Baak

Second Supervisor:
Dr. Peter Kluit

Second Reviewer:
Prof. Paul Kooijman
1 Introduction

2 Higgs mechanism-theoretical background
   2.1 Higgs decay
      2.1.1 Predicted cross sections
      2.1.2 Scattering of the $WW \to WW$
   2.2 Vector Boson Fusion
      2.2.1 Phenomenology – Experimental
      2.2.2 Higgs transverse mass
      2.2.3 $W$ pair spin correlation

3 LHC and the ATLAS experiment
   3.1 LHC
   3.2 The Atlas Experiment
      3.2.1 Coordinate system
      3.2.2 The inner detector
      3.2.3 Calorimetry
      3.2.4 Muon spectrometer
      3.2.5 The trigger system
   3.3 Monte Carlo Samples
   3.4 Reconstruction
      3.4.1 Particle reconstruction

4 Software frameworks
   4.1 Atlas Modular Analysis
   4.2 Toolkit for Multi-Variate Analysis
      4.2.1 Rectangular cut optimization (RCO)
The Standard Model is the framework that justify how particles interacts with the forces that govern nature. It is a successful theory and during the last 60 years it has been proved. The only missing piece in the theory is the Higgs Particle and this is the particle that gives other particles mass. Therefore, the high energy physics community is searching intensively to identify the existence of the Higgs particle, or at least to discard its existence in as many bands of mass as possible.

In an effort to give an experimental answer to this and other questions, the Large Hadron Collider (LHC) has been created, with a high center of mass energy of collisions ($\sqrt{s} = 14$ TeV) and a high design luminosity of $10^{34} cm^{-2}s^{-1}$. ATLAS is the largest detector at the LHC. Its main purpose is the detection of the aforementioned Higgs particle.

Since the Higgs mass is unknown, many mass channels need to be studied. This work studied the Higgs boson produced through vector boson fusion with a final decay into two W boson, in an expected energy band of 170 GeV.

This thesis first presents a theoretical introduction to the Higgs produced through Vector Boson Fusion channel and a motivation for its study when the Higgs boson decays into two W. Then it briefly discusses some design and performance requirements of the ATLAS detector in the LHC and the next chapter explains the software programs and the multivariate analysis tool used in this work. Then my pre-selection cuts to the Monte Carlo samples are discussed in detail as well as the tuning of the multivariate methods selected to optimize the sensitivity for finding the Higgs boson in the selected channel. Finally the optimization results for different methods are discussed.
In nature there are 4 known fundamental ways in which particles interact with each other: the strong, electromagnetic, weak and gravitational interactions. Those are the forces that govern the interaction between the particles that constitute matter. The framework that explains both the interactions and the composition of matter is called the Standard Model (SM)[17].

This model describes all fundamental particles and their behavior under the three strongest forces, the strong, electromagnetic and weak interaction. Currently, no theory has successfully incorporated gravity into the SM.

According to the SM, matter is composed of fermions, which are spin $-\frac{1}{2}$, point-like particles, and they are divided in 3 families of leptons ($l=\{e, \nu_e\}, \{\mu, \nu_\mu\}, \{\tau, \nu_\tau\}$) and quarks ($q=\text{u,d,c,s,t,b}$) (Table 2.1). The Standard Model is based on the gauge principles, which explain that the forces are mediated through Gauge Fields[24]. The particles responsible for mediating the forces are gluons, $W^\pm$ and $Z^0$ gauge bosons and are point-like particles with spin $-1$ (Table 2.2).

The reason some particles have mass and others do not is not explained in the Standard Model. The answer to it may lie in the mechanism of spontaneous symmetry breaking[19], based on the existence of Gauge Fields. The mechanism leads to an additional new particle called the Higgs boson spin-0. That is the responsible particle for giving mass to all other particles. The Higgs boson is the only particle from the Standard Model that has not been experimentally observed. For the interested reader, the following books and papers contain detail explanation of the Higgs mechanism [26][19][24][28].

This chapter will deal with the possibilities of finding the missing Higgs gauge boson with the ATLAS experiment, the characteristics of the decay
2. Higgs mechanism— theoretical background

### Table 2.1: Fermions divided in 3 families of leptons and quarks [30].

<table>
<thead>
<tr>
<th>Leptons spin $\frac{1}{2}$</th>
<th>Mass $(MeV/c^2)$</th>
<th>electric charge</th>
</tr>
</thead>
<tbody>
<tr>
<td>e (electron)</td>
<td>0.511</td>
<td>-1</td>
</tr>
<tr>
<td>$\nu_e$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\mu$ (muon)</td>
<td>105.658</td>
<td>-1</td>
</tr>
<tr>
<td>$\nu_\mu$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\tau$ (tau)</td>
<td>1776.84</td>
<td>-1</td>
</tr>
<tr>
<td>$\nu_\tau$</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Quarks spin $\frac{1}{2}$</th>
<th>Aprox mass (MeV)</th>
<th>electric charge</th>
</tr>
</thead>
<tbody>
<tr>
<td>u</td>
<td>2.55</td>
<td>$\frac{2}{3}$</td>
</tr>
<tr>
<td>d</td>
<td>5.04</td>
<td>$\frac{1}{3}$</td>
</tr>
<tr>
<td>c</td>
<td>1270</td>
<td>$\frac{2}{3}$</td>
</tr>
<tr>
<td>s</td>
<td>104</td>
<td>$\frac{1}{3}$</td>
</tr>
<tr>
<td>t</td>
<td>171200</td>
<td>$\frac{2}{3}$</td>
</tr>
<tr>
<td>b</td>
<td>4200</td>
<td>$\frac{1}{3}$</td>
</tr>
</tbody>
</table>

### Table 2.2: Vector Boson mediators of the particle interactions of the Standard Model, according to the Particle data Group (PDG) [30].

<table>
<thead>
<tr>
<th>Interaction</th>
<th>Boson</th>
<th>Mass $(GeV/c^2)$</th>
<th>Spin</th>
<th>Charge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electromagnetic</td>
<td>$\gamma$</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Weak</td>
<td>$W^\pm$</td>
<td>80.403 ± 0.029</td>
<td>1</td>
<td>±</td>
</tr>
<tr>
<td></td>
<td>$Z^0$</td>
<td>91.1876 ± 0.0021</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Strong</td>
<td>$g$</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

channel of interest for this work, and the motivation behind the selection of this channel.

#### 2.1 Higgs decay

The Standard Model Higgs has not yet been detected experimentally, and its mass is free in the theory. Direct searches at Large Positron Electron collider (LEP) have set an exclusion limit at 114.4 GeV (Fig. 2.1). Precision electroweak measurements show that the mass should be less than 154 GeV, at 95% confidence level (blue band, including the theoretical and experimental uncertainty).

ATLAS will search for the decay of the Higgs boson in various final states. These channels are (Fig. 2.2):
Fig. 2.1: Theory uncertainty plot. The lower limit on the Higgs mass is 114 GeV and the Upper limit is 154 GeV, with 95% CI (blue band).

- $H \rightarrow \gamma \gamma$ in direct production;
- $H \rightarrow \gamma \gamma$ in associated production, using the signatures of the associated particles $WH$, $ZH$ and $t\bar{t}H$;
- $H \rightarrow b\bar{b}$ in associated production of $WH$, $ZH$ and $t\bar{t}H$, using b-tagging;
- $H \rightarrow ZZ^* \rightarrow 4q$;
- $H \rightarrow ZZ$ and $H \rightarrow ZZ \rightarrow llqq$;
- $H \rightarrow WW \rightarrow l\nu jj$ and $H \rightarrow ZZ \rightarrow llqq$;
- $H \rightarrow WW^* \rightarrow l\nu l\nu$;

The two main mechanisms to produce a Higgs in the ATLAS experiment are gluon-gluon fusion and Vector Boson Fusion (VBF). In the first one, two gluons couple to a heavy fermion loop and the Higgs is emitted from this
2. Higgs mechanism-theoretical background

The second is the Vector boson Fusion mechanism, where incident quarks emit virtual $W'$s and $Z'$s, that collide to form the Higgs boson, see Fig. 2.4. An extended explanation and motivation for the studying of Vector Boson fusion process will follow.

![Signal significance graph](image)

Fig. 2.2: ATLAS sensitivity for the discovery of Standard Model Higgs. The statistical significance has been computed as $S/\sqrt{B}$. The Blue squares in the plot correspond to the signal studied in this work, $H \rightarrow WW^* \rightarrow l\nu l\nu$, where $l$ is a lepton and $\nu$ a neutrino.

### 2.1.1 Predicted cross sections

From the different channels with vector bosons in the final state, the one selected in this thesis is $H \rightarrow WW \rightarrow ll\nu\nu$, produced in $qq$ collisions. For a Higgs mass of 170 GeV this channel has an appreciable production cross section and large branching ratios into leptonic final states. Though this is not the channel with the highest rate it does have a clear signature, and backgrounds can be suppressed using techniques that will be explained later.

Table 2.3 shows that the highest rate for this channel occurs for a Higgs mass of 170 GeV. The Higgs mass can be deduced from the WW-invariant
2.1 Higgs decay

![Feynman diagram of the gluon-gluon fusion decay channel.](image)

Fig. 2.3: Feynman diagram of the gluon-gluon fusion decay channel.

<table>
<thead>
<tr>
<th>$m_H$ (GeV)</th>
<th>130</th>
<th>140</th>
<th>150</th>
<th>160</th>
<th>170</th>
<th>180</th>
<th>190</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma$ (qqH) pb</td>
<td>4.04</td>
<td>3.72</td>
<td>3.46</td>
<td>3.22</td>
<td>3.06</td>
<td>2.82</td>
<td>2.64</td>
</tr>
<tr>
<td>$\sigma$.BR($H \to WW$) fb</td>
<td>1127</td>
<td>1785</td>
<td>2370</td>
<td>2955</td>
<td>2969</td>
<td>2620</td>
<td>2054</td>
</tr>
<tr>
<td>$\sigma$.BR($H \to \tau\tau$) fb</td>
<td>223</td>
<td>135</td>
<td>64.4</td>
<td>11.9</td>
<td>2.8</td>
<td>1.6</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Table 2.3: Vector boson fusion production cross-section $\sigma$ (qqH) and $\sigma$.BR($H \to WW$) as a function of the Higgs mass [27].

The mass of the W itself has been measured to be $M_W = 80.398 \pm 0.025$ GeV [30], which results in the maximum production rate at about $2M_W$.

2.1.2 Scattering of the $WW \to WW$

The LEP2 experiment (Large Electron Positron Collider), produced evidence of the first clear proof of the presence of WW couplings, and the W-pair sample allowed the first direct measurement of the W hadronic and leptonic decay branching ratios [22].

One of the motivations for studying Higgs decaying into two W’s is because the scattering of $WW \to WW$ is a good place to confirm the Higgs boson or reveal new physics beyond the Standard Model. In this process, the Higgs particle needs to be found in order to guarantee the unitarity and renormalizability at the $WW \to WW$. If no Higgs boson exist, something else must show up in the scattering of the $WW$ to ensure that the total amplitude stays renormalizable. As such the scattering of $WW \to WW$ is an excellent prove of the electroweak symmetry-breaking sector. Therefore this study will lead to the “no lose” theorem. For a detailed explanation see Ref. [18] and [21].
2.2 Vector Boson Fusion

The Vector Boson Fusion process (Fig. 2.4), is the second largest cross-section for the Higgs and comes from the fusion of vector bosons radiated from initial-state quarks. In this process the Higgs boson is centrally produced from two heavy boson $W^+W^-$. The two quarks, enter the forward region of the detector produce the so-called tagging jets. The W’s can be indirectly measured detecting their leptonic decays.

From different studies [9] it is known that if the decay of the W’s is hadronic, the detection of the boson would be too difficult because the signal would be overwhelmed by the background [18]. This is why the best option is the W’s decaying into double lepton eg. $W^\pm \rightarrow l^\pm \nu$.

![Fig. 2.4: Feynman diagram of Vector Boson Fusion decay channel.](image)

2.2.1 Phenomenology – Experimental

This is an EW production process process with two highly energetic jets in the forward directions of the detector. The Higgs signal has the following properties, as can be see in Fig. 2.4.

- Absence of central jets: the lack of color exchange between the initial-state quarks and the decay product of the Higgs leads to suppressed hadron production in the central region. In this way, it is possible to have a veto on additional jets in the central region [27], [7]. This suppresses dramatically any backgrounds that have color exchange and therefore jets in the central region.

- Central leptons: there are very energetic leptons coming from the Higgs decay in the central region, with opposite charges, traveling preferentially in the same direction. This happens because spins of the two W are anti-correlated (see next section).

- Two highly energetic jets at high rapidity: Two jets (jets are explained in 3.4.1) come from the initial state quarks that emit the heavy vector
bosons. Typically they have high $p_T$ and lie in different hemispheres, i.e. in the forward and backward region of the detector.

- Missing energy: The VBF events have missing $E_T$, because of the undetected neutrinos.
- Absence of b-jets

In conclusion, we will be looking for events with little hadronic activity in the central region, with a positive and negatively charged lepton leaving the interaction region in the same direction. This accompanied by a large transverse momentum jets in both the forward and rear directions.

### 2.2.2 Higgs transverse mass

The Higgs mass peak can not be reconstructed directly from the VBF $H \to WW$ channel. So, to calculate the Higgs mass the distribution of the transverse mass is used:

$$m_T^2 = 2m_{ll}^2 + 2\sqrt{m_{ll}^2 + p_T^2 \sqrt{m_{ll}^2 + p_T^2} - 2p_T^2 \cos \beta}$$  \hspace{1cm} (2.2.1)

For this formula, it should be noted that: $\beta$ is the angle between 2 leptons. $p_T$ is the transverse momentum of the leptons. The missing energy $E_T^{\text{miss}}$ is used as the neutrino momentum $p_T^\nu$. The Higgs is assumed to be produced at rest, so the mass of the leptons is approximate to the mass of the neutrinos $m_{ll}^2 = m_{\nu\nu}^2$.

### 2.2.3 W pair spin correlation

The Higgs boson is supposed to have spin-0, while the $W^\pm$ boson has spin-1. For VBF to conserve angular momentum, it is necessary for the W’s spin to be anti-correlated. Let’s consider the Higgs rest frame. In this frame the W are quantized $S_3(W) = \pm 1, 0$. Defining transverse (T) and longitudinal (L) polarization, the decays that are allowed are:

$$H \to W_T^+W_T^-, H \to W_L^+W_L^-$$  \hspace{1cm} (2.2.2)

It is not possible to measure the polarization of the $W^\pm$ directly, however, it is possible to measure the final state charged leptons. In the SM, the W’s produced by $H \to WW$ are preferentially longitudinally polarized.
In the decay $W^- \rightarrow e^-\bar{\nu}_e$, the neutrino emitted by the $W^-$ is right handed. This means that the electron has to travel in the direction opposite to the $W^-$ polarization. On the other hand, for $W^+ \rightarrow e^+\nu_e$ the opposite happens: the neutrino is left handed, and the positron follows the direction of the $W^+$ polarization \[26\] \[6\]. Since the W-spins are anti-correlated, it follows that both leptons travel preferentially in the same direction, see Fig 2.5.
The Large Hadron Collider (LHC) is a super-conducting accelerator located at CERN (Conseil Europeén pour la Recherche Nucléaire), in Geneva, Switzerland.

The experiment aims to complete and possibly extend the Standard Model of elementary particles. Experimentation at LHC will start at the end of the summer of 2008. Two proton beams moving in opposite directions will be accelerated in a ring of 27 km diameter, placed 100 meters underground. The beams consist of compact bunches of $10^{11}$ protons, 25 ns apart, leading ultimately to a luminosity of $10^{34} \, \text{cm}^{-2} \, \text{s}^{-1}$, providing proton-proton collisions with a center of mass energy of 14 TeV.

The pp collisions will take place at four points along the ring, and around each one, an experiment is located. In one of these points is the ATLAS experiment. Its main goal is to search for the Higgs boson, but it is also searching for new physic phenomena like supersymmetry.

### 3.1 LHC

At the LHC two beams of hadrons (protons or lead ions) will travel in opposite directions inside the accelerator. The LHC is design with a center of mass energy of 14 TeV, with 7 TeV per beam, but will reach that value after a series of steps.

Stage A: The first collisions will last 3 months, during which the tuning of the machine will take place. After about 50 days of operations a peak luminosity of $1.10^{31} \, \text{cm}^{-2} \, \text{s}^{-1}$ is expected.

The stage B will last 6 weeks, during which time the intensity of the
beams is increased, and eventually reaches a luminosity of $5 \times 10^{32}$.

The last step during 2008 is the stage C. This stage is to obtain a maximum luminosity of $10^{33} \text{ cm}^{-2}\text{s}^{-1}$. For center of mass energy of 10 TeV will be reached before the winter shut-down.

From 2009 onwards, as long as the magnet’s performance allows it the LHC, is foreseen to rise the center of mass energy gradually to 14 TeV [1].

3.2 The Atlas Experiment

A Toroidal LHC ApparatuS (ATLAS) is one of the multi-purpose experiments at LHC. Its dimensions are 44m in length, 25m in height and 25m in width. It is composed mainly of three detectors parts: the inner detector (ID), the electromagnetic and hadronic calorimeters, and the muon spectrometers (Fig.3.1). The ID is a tracking detector and measures the trajectories of charged particles. It is in the core of the experiment, surrounding the interaction point, and is itself surrounded by a solenoidal magnet which generates a 2 Tesla magnetic field parallel to the beam axis. By measuring the curvature of the trajectory of the charged particles produced in the collisions, it is possible to determine their momentum.

The electromagnetic and hadronic calorimeters surround the ID and will measure the energy deposits of neutral and charged particles. Finally, the outer most layer is comprised of muon spectrometers. It has high precision tracking chambers that measure the deflection of muons due to the magnetic fields of three toroidal magnets (forward, rear and barrels), having a field of 0.5 Tesla.

3.2.1 Coordinate system

The Atlas coordinate system upon which all measurements are referenced, is defined as follows. The z-axis lays is defined along the beam direction, the positive x-axis is oriented in the horizontal plane toward the center of the LHC ring, and the positive y-axis is pointing vertically upward. The azimuthal angle $\phi$ is in the x-y plane and ranges $[0, 2\pi]$. The polar angle $\theta$ starts at the positive z axis and ranges $[0, \pi]$. The radius is defined as $r = \sqrt{x^2 + y^2}$. The pseudo-rapidity is defined as $\eta \equiv -\ln(\frac{\theta}{2})$. Last, the pseudorapidity-azimuthal space is defined as $\Delta R = \sqrt{\Delta \eta^2 + \Delta \phi^2}$.

3.2.2 The inner detector

The inner detector’s function is to reconstruct the trajectories of charged particles in an environment with a high density of tracks. In order to do
so, the inner detector is placed in a 2 T solenoidal field and consists of three different sub-detectors: the pixel detector, the semiconductor tracker and the transition radiation tracker.

The pixel detector is located around the beam pipe, and is comprised of silicon pixels. The total system consists of 80 millions pixels, providing high granularity.

The semiconductor tracker (SCT), consists of four double layers of silicon strips, surrounding the pixel detector, providing accurate measurements in the r-\( \phi \) and Z plane. It is able to measure the radial position of the hit.

The transition radiation tracker (TRT) is a straw detector. The drift straws have a diameter of 4 mm and a maximum length of 150 cm. This sub-detector is able to detect the transition radiation occurring when a relativistic particle crosses the boundary between two media with different electrical properties. The transition radiation tracker gives information on the particles identity; the lower limit of the velocity of a particle is determined and this, together with its momentum provides and upper limit for the mass of the particle. As such, this measurement helps separate electrons from pions.
3.2.3 Calorimetry

The ATLAS calorimeters consist of an electromagnetic calorimeter (ECAL) and a hadronic calorimeter (HCAL). They are used to measure the energy of most of the particles produced in the interactions. The electromagnetic calorimeter (the inner part) should absorb and measure most (or all) of the energy of electrons and photons, while jets produced by hadrons lose only a small fraction of their energy in this calorimeter. The hadronic calorimeters, which is built around the ECAL, absorbs and measure the remaining energy of the jets.

Also for the particles that can not be “seen” by the detectors (i.e. neutrinos) the calorimeters can “measure” them. Before the collision, the protons only have momentum in the z-direction and the total momentum in x-y plane is zero. Then, after the collision when the momentum of all particles and jets are added, because of conservation of momentum, it is possible to calculate the missing momentum in the transverse plane.

The $E_{\text{miss}}^t$ resolution has been determined for the Atlas detector [5], as:

$$\sigma_{\text{miss}} = 0.46 \sqrt{\sum E_{\text{miss}}^t \text{GeV}}$$ (3.2.1)

The overall system provides very good jet and $E_{\text{miss}}^t$ performance.

The calorimeters consist of two types of materials, absorbers and active materials. A purely active material calorimeter would be too expensive. So an absorber material is placed between the active material to limit the depth of the particle showers.

For the electromagnetic calorimeter lead is used as an absorber and liquid argon as the active material. The overall sub-detector covers a range of $|\eta| < 3.2$ [5]. The calorimeter is divided into a barrel section ($|\eta| < 1.475$) and two end-caps ($1.375 < |\eta| < 3.2$). The thickness of the barrel section is more than 24 radiation lengths [6] and for the end-caps is more than 26 radiation length.

The hadronic calorimeter is divided in three sub-detectors. The barrel calorimeter ($|\eta| < 1.7$) uses iron plates as absorbers and plastic scintillator tiles as active material. It has a total thickness of 11 interaction lengths [7]. The hadronic end-cap ($|\eta| > 1.5$) uses liquid argon. Finally, there is a forward sub-detector, that provides both electromagnetic and hadronic energy measurements. It is a copper-tungsten scintillator calorimeter and it extends the coverage to $|\eta| = 4.9$.

---

1The radiation length is defined as the distance over which a high-energy electron on average loses all but $1/e$ of its energy through bremsstrahlung. [20]

2The hadronic interaction length is defined as the mean free path of a high-energy hadron before undergoing an inelastic interaction.
3.2.4 Muon spectrometer

The muon spectrometer covers the largest volume of ATLAS. It is the outermost layer of the detector. The only particles expected to pass the calorimeters are muons and weakly interacting particles like neutrinos. So the muon spectrometer have been used for measuring high momentum muons coming from e.g. Higgs decays. The two main purposes of this sub-detector are to have an independent muon trigger and high quality stand alone tracking. This means that the muon detector does not need any information from the other sub-detectors to reconstruct a muon.

An important feature of the sub-detector is that it has its own magnetic field independent of the inner detector. It is provided by an air-core toroid magnet with an average field strength of 0.5 T. To achieve its goals the muon spectrometer consist of four types of detection chambers:

The first are the Monitored Drift Tube (MDT) chambers, with which the muon are tracked and their momenta determined. They are made of aluminum tubes filled with $\text{Ar:CO}_2$ and a central wire. When the muons cross the chambers, the gas is ionized and the clusters of electrons produced will drift toward the wire. The Cathode Strip Chambers (CSCs) in the inner forward region have fine granularity and are faster in operation. Finally, the Resistive Plate Chambers (RPCs) and Thin Gap Chambers (TGCs) provide the trigger.

3.2.5 The trigger system

Owing to the high luminosity of the LHC a tremendous amount of data will be produced. Combining the information of all sub-detectors, it is estimated that the data rate produced by Atlas will be around 40 TB/s. This is much higher available than the currently processing speed and too much data to store.

The trigger system has been designed to filter the data stream to a manageable number of events, without loosing interesting events in the pre-selection. The system is organized in three levels that reduce the data by a factor of $10^6$, i.e. from the expected 40 TB/s to 200 MB/s.

The level-1 trigger (LVL1) is a hardware based trigger. Because of the high event rate, the latency or decision time is only 2 $\mu$s. This time constraint allows one to use only basic information of the particles ( i.e. $p_T$ of the particle), that is obtained from the calorimeters and the muon trigger chambers. When the events are accepted, the information from the other sub-detectors is stored.

The level-2 trigger (LVL-2) is based on software. The latency of this trig-
The region of interest determined by LVL1 are processed. If an event is rejected by the LVL2 the data is deleted. Otherwise, the event will pass to the Event Filter (EF) for more advanced analysis.

The Event Filter is the last level of the trigger system. The EF accesses all the information with full granularity. It takes a few seconds to use detailed reconstruction algorithms. The LVL3 runs in a computer farm located close to the detector.

The events that pass the EF trigger are written into mass storage devices and are available for further analysis with the off-line software.

3.3 Monte Carlo Samples

Several event generators have been developed to simulate p-p collisions at 14 TeV data. These programs simulate the particles produced in the collision, their four momenta, and their decays chains. In this work, the event generators used have been Pythia \[15\] for the Higgs product, Herwig \[12\] for the backgrounds and MC@NLO for $t\bar{t}$. The MC samples produced were VBF $H \rightarrow WW$, $gg \rightarrow H$ and for background studies the follow channels were generated $t\bar{t}$, $Z + njets, W + njets$ and $WW$.

Simulation

The detector simulation Geant4, simulates the ATLAS detector \[14\]. It considers the interactions of the particles with the detector material and the passage of them through the magnetic fields.

Response

The detector response and electronics are simulated in the digitization phase. The digitization in the silicon strip detector is done in the following way: after in the detector simulator the free charge produced by the energy deposits drift to the strips in the sensor, the program simulates the read out of the strips and the electronics’ response.

Fast Simulation

The simulation of the detector and detector response is a slow process. A short cut for the reconstruction is the program called ATLASFAST \[11\]. Fig. 3.2. It replaces the full detector simulation and reconstruction phases of the reconstruction chain. The parameters of the detector can be tuned to the expected value of the performance of ATLAS from full simulation. ATLASFAST combines the detector response and detector simulation into one step.

Though this simulation is less precise, it reduces the time of analysis from half an hour to two minutes approximately.
3.4 Reconstruction

Reconstruction is the process where the raw data, like pulse heights and readout times from the detectors, are analyzed and reconstructed into tracks and energy deposits to detect and recognize particles interesting for us [23].

To do so, the reconstruction algorithms perform among others pattern recognition, track fitting, vertex determination and energy measurements. The next step is to take the reconstructed events and compare them with simulated input “truth events”, so that it is possible to study the quality of the reconstruction. Afterward, the output is written as “ESD” event summary data or “AOD” analysis object data. The ESD data contains detailed data of the reconstruction (complete information of one track, with times and energy), while the AOD keeps only the information for physics analysis, like the particle with the four momenta.

3.4.1 Particle reconstruction

Leptons

The leptons used in this work were electrons and muons. In order to consider leptons isolated, the sum of transverse energies of all tracks from the same vertex as the lepton in a cone of $0.01 < \Delta R < 0.2$ around the lepton must be below 5 GeV [3].
3. LHC and the ATLAS experiment

**Electrons**

An electron is a negatively charged particle with light mass \(0.511 \text{MeV}\). It will produce a track and its energy is mostly absorbed in the electromagnetic calorimeter. The momentum of the particle is calculated from the track curvature inside the inner detector, while the energy is measured from the energy observed in the calorimeter.

A jet (discussed below) contain many particles, including many electrons. Sometimes a jet is misidentified as an electron. In order to recognized and electron produced in a jet from one coming from other process, an isolation cut is used. This implementation assures no near by objects in the \(\phi-\theta\) space.

**Muons**

Muons are charged particles heavier than electrons \((105.659 \text{ MeV})\). They leave a track in the inner detector but they penetrate the calorimeter and deposit only a small amount of energy in the calorimeter. They leave a track in the muon spectrometer.

**Quark and gluon jets**

A jet is a cone of hadrons and other particles. It is produced by hadronization of a quark or a gluon. In this process a large number of individual particles are produced. The charged particles will leave a track in the inner detector, but most of the particles are expected to deposit most of their energy in the electro-magnetic and hadronic calorimeters.

**b-jets**

B-mesons have a relatively long lifetime and therefore originate from a secondary vertex. Therefore, a b-jet and can be identified with a secondary vertex reconstruction.
Within the ATLAS experiment, software simulation and analysis framework has been developed for the reconstruction algorithms and physics analysis. This chapter will explain briefly the simulation programs with Monte Carlo Higgs event simulations as well as the development of Atlas Modular Analysis (AMA), an analysis program developed at NIKHEF. Finally, the Toolkit for Multivariate Analysis (TMVA), the principal tool used in this work for the optimization of the Higgs signal observables will be introduced. By applying TMVA this work is after finding the right observables to split the signal from background.

4.1 Atlas Modular Analysis

Atlas Modular Analysis (AMA) is a module based program developed at NIKHEF to facilitate the analysis of mainly root based (Root is the object oriented framework for large scale data analysis developed at CERN [16]) n-tuples made from ATLAS analysis object data (AOD).

The idea is to analyze the n-tuples in a simple way. The software has different directories (framework, modules, samples, summary) that allows to introduces technical information, to perform tasks such as filtering, event and particle selection with the introduction of cuts, which can easily be changed or turned off depending on the analysis, and the possibility of private analysis modules are easily implemented. Each event is automatically read in, and the individual particles stored as particles four-vector [29].
4.2 Toolkit for Multi-Variate Analysis

In ATLAS there is a need for searching for a small Higgs signal in a large data set. It is essential to extract the maximum of available information from the signal characteristics. TMVA has been designed to find the best separating function between the signal and background. It contains a variety of multivariate classification algorithms \[10\]. For this work the following methods have been selected: rectangular cut optimization, Fisher method (linear cut), artificial neural network and boosted decision tree (non-linear cut), see Fig. 4.1.

The algorithms consist of two phases. The first one is the training, where the program learns to classify data from a finite but representative set of samples. In the second phase, the already trained classification system is tested against new samples unknown to it. In this way it is possible to assess its real classification capabilities for arbitrary samples of data.

![Types of cuts in multivariates methods: a) linear cut b) rectangular cut c) non-linear cut.](image)

Fig. 4.1: Types of cuts in multivariate methods: a) linear cut b) rectangular cut c) non-linear cut.

4.2.1 Rectangular cut optimization (RCO)

Being the simplest of TMVA methods, the Rectangular Cut Optimization seeks for the extraction of signal events from background data by applying rectangular cuts on selected variables. A binary decision tree is created by filtering different variables between a minimum and maximum value, and the performance of the classifier depends on the joint effect of each one of the cuts in the chosen combination.

To optimize the cuts assembly, this work uses the genetics algorithms mode. In it, a first set (or generation) of cut assemblies is generated randomly, and an index is used to assess and compare their performance. The best sets are taken apart and their characteristics are used as a seed to generate a
new generation. This process repeats itself in a manner similar to genetic evolution, and yields generations of cuts highly adapted to solve the problem under study.

Due to its simplicity, the methods performance greatly degrades with the discriminant power of the inputs. Hence, it is advisable to use RCO with inputs with large discrimination power.

4.2.2 Fisher discriminants

This method is based on differentiating the categories by making first a linear combination of the different features or dimensions of the data. This yields a projection of that data on a single axis, where different ranges or vicinities can be associated to each category. During the training phase, linear combinations are optimized by maximize the distance between the means of the different classes (signal and background) and minimizing the dispersion in each class.

Once the axis is selected, the projection of the samples on the axis are averaged and a position in the axis is determined that best represents the category. New samples are compared to the possible categories, and the closest one is then chosen. This method provides a simple and fast way to classify samples in spaces with high dimensionality and dispersion of the data, specially with linearly correlated Gaussian-distributed variables. On the other hand, it is of no use for data where the classes share their mean values, regardless of their distribution shape.

4.2.3 Boosted decision trees

A decision tree is composed of joints and branches. The data arrives from the trunk (principal entry of the data) to the first node where a decision is made based on an specific characteristic of the data. Depending on the result of that decision, the data takes one of two possible branches to the following joints and eventually reaches terminal points called leafs.

These are labeled with the different categories of the data, answering, in our case, whether the data that arrived to them corresponds to signal or background. Boosted decision trees (BDT) can be seen as a forest of decision trees. This multivariate classifier weights the majority of vote of each of the individual trees that compose the BDT.

Due to the binary split method, traditional decision trees are not optimal when the variables used to characterize the data are linearly correlated. There exists however an improved version called De-correlated BDTD which minimizes the negative effects of correlation, and which is used.
4.2.4 Artificial neural networks (ANN)

ANN stands for a network of artificial neurons. In this context, a neuron is an element that takes a series of inputs, makes a weighted sum of them, passes the result through an activation function and gives an output from it. A network of such elements generates a nonlinear mapping from inputs to outputs and given the existence of a method to select the appropriate weights, scaling factors and functions, that mapping can represent an arbitrary function. And in particular, it can represent the classifier we need, where a space of many observables need to be mapped onto a one-dimensional response indicating whether the data is more likely to corresponds to signal (1) or background (0).

In this work, multi-layer-perceptron (MLP) neural networks are used. They are composed of a first layer of neurons receiving directly the input signals, one or more layers of hidden networks that combine and analyze the data, and a neurons layer directly connected to the output ports. In the training phase, a back propagation schema evolves the weights of the connections in order to maximize the efficiency of extraction of the signal from the background. In a iterative process, after presenting an example and checking the answer given by the network, each weight is adjusted by a factor proportional to the magnitude of the output error and to the responsibility that that specific connection had in producing it. After many steps the network evolves and refines its mapping into the desired one. Last but not least Fisher method is the most robust against over-training.

4.2.5 Over-training

A problem usually encountered in all the methods, but mostly in BDT and ANN, is over-training. It occurs when the size of the decision system is too big for the complexity of the training data, when the size of the training set is too small, or when it is not representative of the whole input space. In such cases, the classification method specializes itself in differentiating the specific examples given for training, learning them by memory instead of learning the underlying rules and principles necessary to classify them. In such cases, an artificial and exaggerated discrimination power is observed when the system is applied to the training data. However, since the inference capabilities of the system are in fact not developed, the system shows a poor performance when analyzing data unknown to it.

In order to avoid this problem, it is important to use large enough training sets, from all over the input space. Also in some cases, limiting the size and complexity of the systems force them to focus more attention on the general
characteristics and differences of the studied categories. Finally, for assessing the real capabilities of the already trained system, and for detecting any possible over-training issues, a sizable and representative test-set is essential.

The RCO and Fisher methods are more robust against over-training. In the case of RCO, the method has a very continuous behavior within each side of the selected frontiers. So, if the training set clearly defines the average borders of the interest zone, all the training and testing samples within it will be treated in the same way and accepted. For Fisher, a similar situation maybe helping to avoid the over-training. The linearity of the method and simplicity of the system force it to work on the average characteristics of the data, again yielding to an homogeneous response to training and testing samples (given a representative enough training set). On the other side, the same complexity possibilities that give BDT and ANN their classification power, may allow a too big system to specialize in the specific training points. Meanwhile, and given their non linearity, the homogeneity and continuity of their response is sacrificed, yielding to different behaviors between the training points and their neighboring testing ones.
CHAPTER 5

HIGGS PRE-SELECTION

At the LHC millions of proton-proton collisions will occur per second, and millions of particles will be created. Even if the Higgs boson exists, there will be only few hundreds of them. So, how will it be possible to detect?

The ATLAS experiment needs to carefully select the Higgs boson signal from the copious p-p collision events. The first part of the selection is the triggering process (section 3.2.5), where it is decided on-line which events are interesting and should be kept for off-line analysis. From the large number of saved events, one needs to choose the channels that are interesting to study. The goal of the pre-selection is to try to select efficiently the Higgs events, while rejecting background events. The pre-selection is a set of cuts based on the characteristics of the signal events.

The next step is to understand what the signal that is to be measured, looks like, and what are their characteristics. In section 5.1, the characteristics of Vector Boson Fusion Higgs productions are discussed. The cross section and characteristics for various possible backgrounds are addressed in section 5.2. Monte Carlo samples and pre-selection criteria are presented in section 5.3 and 5.4. Finally, a summary of the observables, used to characterize the signature are shown in section 5.5.

5.1 Signal characteristics

In the VBF channel, $qqH \rightarrow W^{(*)}W^{(*)}$, the Higgs is produced centrally. Figure 5.1 shows the Feynman diagram. The W’s decay into 2 leptons and their neutrinos, $e, \mu, \nu_e, \nu_\mu$. The original quarks, which emitted the vector bosons enter in the detector producing jets in the forward region.
5. Higgs pre-selection

Fig. 5.1: Feynman diagram of the Vector Boson Fusion decay channel. The Higgs decay into two W’s and each W decays leptonically.

As discussed in section 2.1, the signal is clean, contrasting with most backgrounds. The jets in the forward region of the detector and the veto of the central jet activity are powerful tools to enhance the signal-background ratio [8]. The leptons are preferentially produced in the same direction, close to each other [6], whereas in the background the directions of the leptons are non-correlated.

In the present study, the VBF Higgs decaying into WW analysis has been performed with MC data with a Higgs mass of $170\, GeV/c^2$ and for $1\, fb^{-1}$, which under nominal operating of the LHC is expected to be accumulated in 100 days of LHC running. This mass is approximately the double of the $M_W$, which would be the most favorable mass for the Higgs in Vector Boson Fusion.

As part of the signal studies, it is necessary to analyze any background channel that produces a similar final state. The backgrounds that have signature of 2 jets + 2 leptons + missing $E_T$ will now be discussed.

5.2 Background characteristics

It is not feasible to produce in MC all possible backgrounds. For this study, it was decided to use all available backgrounds with significant cross sections (see 5.1), and with final states similar to the signal. Those are:

1. $t\bar{t}$
2. $W + n\, jets$
3. $WW$
5.2 Background characteristics

4. $ZZ$

5. $Z + n\text{jets}$

All these backgrounds have at least one high $p_t$ lepton, and two jets as follows:

1. $t\bar{t}$ production: it is the dominant background contribution in the VBF channel. The production cross section is the largest (see Table 5.1). The top decays $t \rightarrow bW$ (Fig. 5.2), and the final decay can be $bbWW$ which leads to a signature similar to the signal $qqWW$. The two $b$’s from the top quark can be identified as the tagging jets (defined in section 5.4). The $b$ jet can be however in any direction of the detector. The W’s are not spin anti-correlated, so the leptons from the W’s mostly have different directions.

![Feynman diagram for $t$ decaying to one $W$ and one $b$ jet.](image)

2. $W + n\text{jets}$ production: The $W$ decays into a lepton and a neutrino, and one of the jets is mis-reconstructed as a lepton. Other jets look like the tagging jets of the Higgs signal. However, the momenta of the leptons do not have preferred direction. There is jet production in the central area.

3. $WW$ production: The mode $WW$ is directly the decay mode of the Higgs channel studied here. Other produced jets are mis-identified as tagging jets.

4. $Z + n\text{jets}$ production: The $Z$ boson can decay to $Z \rightarrow e^+e^-/\mu^+\mu^-$ pair. In this way, $Z$ could decay into leptons and the jets can be confused
with the tagging jets. The main discriminant variable is the missing $E_t$, which is a signature of the Higgs $\rightarrow$ WW decay but not from Z production.

5. ZZ production: There is contribution from one Z decaying leptonically and the other Z decays into hadrons, producing jets. In this process there is no missing $E_t$ either.

5.3 MC Samples used

The Monte Carlo data used in this work were produced with Pythia, Herwig and MC@NLO.

Table 5.1 shows all the MC samples used in this thesis, the average number of events per sample and the generated cross section. The weight value is defined as the necessary scaling that needs to be added to the event to make it equivalent to 1 $fb^{-1}$ of data.

<table>
<thead>
<tr>
<th>MC sample</th>
<th>Average events</th>
<th>Generated cross section (1/fb)</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$VBF$</td>
<td>13250</td>
<td>150</td>
<td>0.0113</td>
</tr>
<tr>
<td>$gg \rightarrow H$</td>
<td>15000</td>
<td>1307</td>
<td>0.0871</td>
</tr>
<tr>
<td>Background</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$tt$</td>
<td>867500</td>
<td>449820</td>
<td>0.5185</td>
</tr>
<tr>
<td>$WW$</td>
<td>186100</td>
<td>1273</td>
<td>0.0068</td>
</tr>
<tr>
<td>$W + nj$</td>
<td>240020</td>
<td>593343</td>
<td>2.4720</td>
</tr>
<tr>
<td>$Z + np$</td>
<td>43050</td>
<td>28762</td>
<td>0.6681</td>
</tr>
</tbody>
</table>

Table 5.1: MC Samples with the average number of events, the cross section and the weight. VBF : $1/fb^{-1} = 150$ Higgs events

The simulated events samples are from the software release 12.0.6, the so-called the CSC sample.

---

1 $fb^{-1}$ is equivalent to having 100 days of running at $10^{32}$ Luminosity
5.4 Pre-selection cuts

Cuts are applied to pre-select the Higgs signal and suppress the background channels.

1. Event selection of the 2 most energetic leptons.

   The leptons need to be isolated, with opposite charges, and with $|\eta| < 2.5$ because the reconstructed track covers the full inner detector. In addition, the leptons must be in between the tagging jets.

   The muons track must have in a radio $\Delta R$ of 0.2 less than 20 GeV ($E_t(20^o) < 20 GeV$). It is asked for a muon track that match the inner detector with the muon-spectrometer. The $p_T > 10$ GeV and $|\eta| < 2.5$ as explained for the leptons.

   To eliminated mis-reconstructed muons were applied extra cuts to the leptons in $\Delta \Phi < 0.2$, $\Delta \eta < 0.1$ and lepton mass $< 6$ GeV.

   The electrons are defined in $\eta = -2.5:-1.57 -1.35:1.35 1.57:2.5$ because the crack areas of the are excluded, the $p_T > 10$ GeV and the isolation cut per track is $E_t(20^o) < 20 GeV$.

2. Selects at most 3 good jets.

   The good jets required to have $|\eta| < 4.8$, $p_T > 40$ GeV. The tagging jets are two good Jets that tend to be well-separated in pseudo-rapidity and the highest-$p_T$ in the event.

   In ATLAS MC data there were 3712 signal events and 11990 background events. After the pre-selection there are 3695 signal events and 11956 background events left.

   *Triggering cuts used*

   In this thesis, the triggering cuts are selected as follows in Table 5.2, to work only with events showing 2 isolated leptons. After using the triggering cuts the Higgs signal sample is reduced in 0.45% and the background sample is reduced 0.4%
Table 5.2: Level1 (L1), Level2 (l2), and EventFilter (EF) and the cuts made by trigger. The first letter stands for the particle, e electron, mu Muon. If a number precede the letter, it is the number of particles required. The number following the letter is the $p_T$ threshold. The i shows and isolated object, i.e no nearby object in $\eta$-$\phi$ space.

5.5 Observables before optimization

This section will cover the selection of the input variables that will help to enhance the separation between the signal and the backgrounds events.

The signal samples used are the gluon-gluon fusion and the VBF both for a Higgs mass of 170 GeV. The background samples events are those specified in 5.2. The selection of the observables is based on using the tagging jets variables, missing energy and the $\Delta \phi$ of the leptons.

- $|\Delta \eta_{jj}|$ pseudorapidity between the tagging jets.
- $|\eta_j|$ Absolute value of the $\eta$ of the jets.
- Transverse mass of the tagging jets.
- Transverse missing Energy.
- $b$ jet weight.
- $|\Delta \phi_{ll}|$ Absolute value of the $\Delta \phi$ of the leptons.
Figure 5.3 shows the separation $|\Delta \eta|$ between the tagging jets. For the Higgs signal the tagging jets are well separated in pseudorapidity, so the peak of the separation must be centralized. For background, there should not be any correlation in the symmetry for the jets.

![Delta Eta vs Counts](image1)

![Delta Eta vs Counts](image2)

**Fig. 5.3:** Left: Separation $|\Delta \eta|$ of the tagging Jets for signal $H_m = 170$ GeV. Right: Separation $|\Delta \eta|$ of the tagging Jets for background events

The missing Et of the signal and the background are shown in Fig. 5.4. As we can see the $E_t^{\text{miss}}$ do not have a big discriminant power to differentiate between the signal and background. This observable is a good example of why it is needed a multivariate optimization program.

![Missing energy vs Counts](image3)

![Missing energy vs Counts](image4)

**Fig. 5.4:** Left: Missing $E_t$ for a Higgs signal events with $H_m = 170$ GeV. Right: Missing $E_t$ for background events.
Fig. 5.5 shows the distribution of the mass of tagging jets for signal and background. The forwards jets comes from the original quarks of the collision, so they are more energetic than the jets identified in the QCD processes.

The $b$ jet weight shows how likely it is that a jet comes from a $b$ quark. This parameter enhances the possibility of eliminating the huge background coming from the $t\bar{t}$. It can be seen in fig. 5.6.

Fig. 5.6: Weight of the $b$ quark. Left: $b$-weight for signal data, the jets do not come from a $b$-quark. Right: $b$-weight for background data, many jets comes from $b$-quark.
In Fig. 5.7 is shown that the Higgs signal has mostly two jets and the background mostly three.

The transverse Higgs mass is shown in Fig. 5.8. For the signal data, the peak is around 170 GeV, as expected. While, for the background data the shape is not related to the Higgs signal.
The $\Delta \phi$ of the leptons exploit the differences in the spin correlations in the W's Fig. 5.9. In the signal, both leptons travel preferentially in the same direction, while in the background this is not the most probable case. The region with a big $\Delta \phi$ is background, and the region with smaller $\Delta \phi$ is signal.

Fig. 5.9: $\Delta \Phi$ of the leptons. Left: $\Delta \Phi$ for Higgs signal. Right: $\Delta \Phi$ for the background. This plot shows a clear difference between the signal and background.

Fig. 5.10 shows $\eta$ for the leptons.

Fig. 5.10: $\eta$ for the leptons. Left: Eta for the Higgs signal. Right: Eta for the background.
CHAPTER 6

OPTIMIZATION OF HIGGS EVENT SELECTION

In order to improve the performance of a multivariate analysis, in general classifiers have tuning parameters to optimize separation between the signal and background candidates. This chapter describes the input observables used in the program as well as the tuning of 3 multivariate methods: rectangular cut optimization (RCO), artificial neural network (Multilayer perceptrons MLP) and boosted decision tree (BDTD). The Fisher method was not optimized, so the method was used as a reference.

Section 6.1 explains the used input variables. The tuned variables for each method used are shown in 6.2. Finally, the sensitivity parameter is presented in section 6.2.4.

6.1 Input parameters selection

The idea of analyzing all possible information coming from an experiment in order not to miss anything may sound tempting. However, the inclusion of trivial or correlated variables can introduce noise into the system without actually providing new information, and thus not improving the performance.

The observables used for the optimization are:

1. Average pseudorapidity $|\eta|$ of the tagging jets.
2. Separation between the forward tagging jets $|\Delta \eta_{jj}|$.
3. Invariant mass of the tagging jets.
4. Missing transverse energy, $E_T^{\text{miss}}$. 35
5. b-weight.

6. $\Delta \phi$ of the leptons.

From these observables, two sets of input variables have been selected, see sections 6.2.1 and 6.2.2.

6.2 Tuning Parameters

The ATLAS MC samples described in section 3.3 have been used to tune the parameters of rectangular cut optimization, multilayer-perceptrons and boosted decision trees.

From the MC samples, 3712 signal events and 11990 backgrounds events are left after the pre-selection. Of these, 85% are randomly selected as training samples, while the remainder are left for testing purposes.

Two characteristic sets of observables have been selected. The first method includes the observables related to the tagging jets and the missing energy. However, in this selection the variables are correlated, see Table 6.1. The second method, in subsection 6.2.2, focuses on uncorrelated variables, see Table 6.3. Observables with an expected good separation of Higgs signal from background like the $\Delta \phi$ of the leptons have been selected for it.

6.2.1 Method 1: Tagging jets variables

$E_T^{\text{miss}}, |\eta|$ and $|\Delta \eta|$ of the tagging jets, mass of the leptons and weight of the b jets.

|                  | jets $|\eta|$ | jets $|\Delta \eta|$ | jets invariant mass | $E_T^{\text{miss}}$ | b-weight |
|------------------|--------------|---------------------|--------------------|---------------------|----------|
| jets $|\eta|$       | 100          | 27                  | 10                 | -16                 | -1       |
| jets $|\Delta \eta|$ | 27           | 100                 | 74                 | -3                  | -14      |
| jets $m_{ET}$    | 10           | 74                  | 100                | 15                  | -8       |
| $E_T^{\text{miss}}$ | -16          | -3                  | 15                 | 100                 | 1        |
| b-weight         | -2           | -14                 | -8                 | 1                   | 100      |

Table 6.1: Matrix correlation of the observables variables in percentage for method 1.

The main characteristic of the signal is the two high energetic tagging jets in the forward parts of the detector. Therefore, as a first approach the input variables involve only the tagging jets and missing energy.
A brief overview of the tuning parameters per method is given, in Table 6.2.

**Tuning parameters for RCO**

For each parameter tuned in this method a cut boundary is set. Each parameter has three options: FMax, FMin and FSmart. FMax disables the maximum cut boundary value, FMin disables the cut of the minimum value and FSmart lets the program choose which cut boundary value should be eliminated.

For the variables related to the separation in $\eta$, it is known that the separation must be greater than zero, so the maximum cut boundary was left open. The same holds for the invariant mass of the jets, there is a pre-selection cut that requires a minimum value for the mass of the jet, so the maximum cut boundary was left open too. For the missing transverse energy and b weight the best values will be obtained with the FSmart option.

**Tuning parameters for multilayer perceptrons**

Three variables have been tuned in this method: NeuronType, Ncycles and HiddenLayers.

The neuron activation function used is a sigmoid which is defined in eq. 6.2.1. This function is used because it allows the system to solve non linear systems and keeps the output constrained.

$$y(x) = \frac{1}{1 - e^x}$$  \hspace{1cm} (6.2.1)
The second one is NCycles, which is the number of training cycles of the neurons. The number of cycles was set to 500. There is an improvement in the significance of the signal, and regularity of the results up to this value. Figure 6.1 shows that after 500 cycles there is no improvement in the performance. The significance \( \text{significance} = \frac{S}{\sqrt{S+B}} \), where \( S \) is the number of signal events and \( B \) is the number of background events) value given by TMVA is the same for NCycles= 500,600,700,1000.

![Graph showing test estimator for the number of cycles for MLP.](image)

Fig. 6.1: Test estimator for the number of cycles for MLP. On the x-axis is the number of cycles.

The third parameter is the number of hidden layers. This parameter defines the network architecture by setting the number of neurons per layer in the network and the number of hidden layers.

The selection of the number of layers was based on the Weierstrass theorem, which states: “For a Multilayer perceptron a single hidden layer is sufficient to approximate a given continuous correlation function to any precision, given an arbitrary large number of neurons in the hidden layer. If the available computing power and size of the training data sample are sufficient, one can thus raise the number of neurons in the hidden layer until the optimal performance is reached” Ref.[10].

The number of selected neurons in one hidden layer has been set to \( N + 5 \), and it has been selected only one layer, where \( N \) is the number of variables (10 neurons). While too small a number of neurons is not effective, an excessive number of neurons and hidden layers slows the process while creating over-training problems (i.e. \( N+100 \)), hence decreasing the performance of the system against the test sample.
Tuning parameters for boosted decision trees

The tuned parameters for BDTD were PruneStrength and nEventsMin. The pruning strength fixes the number of leaves where the tree will end. The idea is to remove the statistically insignificant nodes and minimize the over-training of the trees, Ref.\[10\]. This variable was set to -1. In this way the value is selected by the program yielding an optimum result without a manual try and error process.

The minimal number of events (nEventMin) is the minimum number of events per node in the tree. If this value is too small the risk of over-training rises, and when it is too high (higher than Eq.6.2.2) the efficiency drops.

nEventMin is known for being powerful but unstable, as small changes in this parameter lead to big changes in the result. A standard value from the literature is:

\[
n_{\text{EventMin}} = \frac{\text{training events}}{\text{variables}^2} \cdot \frac{1}{10} \tag{6.2.2}
\]

This together with the number of training samples lead to the value used in this work of nEventMin=45.

6.2.2 Method 2: mixed variables

Tagging Jets \(|\eta|\), missing \(E_t\) and \(|\Delta\phi|\) of leptons.

|                  | jets \(|\eta|\) | lepton \(\Delta\phi\) | jets invariant mass |
|------------------|----------------|----------------------|--------------------|
| jets \(|\eta|\)   | 100            | 1                    | 10                 |
| lepton \(\Delta\phi\) | 1              | 100                  | 3                  |
| jets invariant mass | 10             | 3                    | 100                |

Table 6.3: Matrix correlation of the observables variables in percentage for method 2.

As explained at the beginning of the section, for this method only uncorrelated observables have been selected (Table 6.3). The observable delta \(\phi\) of the leptons has been used, because there is a clear separation between the background and the Higgs signal, as explained in section 5.5.

The tuning of the multi-variates remained almost the same as for tagging jets variables.
### 6. Optimization of Higgs event selection

#### 6.2.3 Over-training

Keeping the system free from over-training is not easy. So many variables affect it that it is difficult to predict its effect on the output during the tuning process. General and empirical rules indicate that keeping the systems simple (few neurons, small trees), and the training process just long enough, forces them to understand the data rather than to learn it, thus avoiding the over-training problem. These principles were followed, but it was still necessary to compare the training and testing set significances every time a parameter was tuned.

#### 6.2.4 Sensitivity parameter

Sensitivity is defined as the amount of signal compared to the error on the background events. TMVA optimizes the signal efficiency $\epsilon_s$ against the background rejection defined as $1 - \epsilon_{Bck}$, where $\epsilon_{Bck}$ is the efficiency of the background. While the tuning of the methods was based on obtaining the highest value for the sensitivity. Nevertheless, both approaches are compatible since if the signal efficiency is high and the background rejection is high the sensitivity will be high as well.

In Punzi [25], it is proved that the best way of computing sensitivity as a means of discovery is:

$$sensitivity = \frac{S}{\sqrt{\sigma/2+B}}$$  \hspace{1cm} (6.2.3)

where $S$ is the number of signal events, $B$ is the number of background events and $\sigma$ corresponds to the number of standard deviations necessary
to distinguish a signal from statistical fluctuation of the background events. For instance a 5 sigma deviation is necessary for claiming the discovery of a new physics phenomena. The Punzi significance protects against optimizing toward zero signal and higher background selection. For example $\frac{S}{B}$ will go to infinity if B=0.
CHAPTER 7

RESULTS OF OPTIMIZATION

After the selection of the methods and the tuning of the parameters, the program was run several times in order to understand its average behavior.

Representative outputs are taken and used to demonstrate and analyze the system capabilities in the following chapter. The performance of TMVA methods for separating Higgs signal from background will be presented. This is done for each of the classifiers separately. After optimization the sensitivity parameter will be shown from which conclusions are drawn on the possibility of detecting the Higgs with the ATLAS detector in the WW decay channel.

In particular, RCO does not combine the different inputs into a single and continuous response. Therefore, for that method the sensitivity parameter is the only way to assess its performance.

7.1 method 1

$E_T^{\text{miss}}$, $|\eta|$ and $|\Delta\eta|$ of the tagging jets, mass of the leptons and weight of the b jets.

After the optimization process of TMVA, the first output generated by the software gives us an overview of the Higgs signal efficiency vs background rejection for the four methods (RCO, Fisher, MLP, BDTD). The full graph is an indicator of the ability to separate “efficiently” the Higgs signal from the background. It is a clear indicator of our system performance and of its capability to fulfill the goal we are optimizing it for.
7. Results of optimization

Fig. 7.1: Signal vs background rejection from TMVA for the four methods optimized (RCO, Fisher, MLP, BDTD)

Fig. 7.1 shows the four methods used. A first glimpse allows to understand which method works best. The background rejection factor is show as a function of the corresponding signal efficiency. A point in the upper right zone of the graph indicates that the method is capable of rejecting most of the background while, at the same time, keeping most of the signal of interest. Therefore, it may seem in the graph that the method with the best overall behavior is the MLP.

Nevertheless, many other output parameters can be studied, and will be analyzed in detail in the following sections. That will allow us to have a more clear understanding of each method’s characteristics, advantages and drawbacks.
Fisher results

Fig. 7.2 shows the results for the Fisher method. This plot shows the count of the output of the system to the Higgs signal input as a dark histogram, and the equivalent for the background as the dash one. The ideal result would be to have signal in one side with a peak around 1 and the background with a peak around zero. That would indicate that the program is able to clearly classify which data corresponds to signal and which one to background. In this case, however, there is not a clear separation between the response to a Higgs signal and to background, and in the entire overlap region, the output of the system would prove useless. The peaks present in the response plots are due to some events with an unusually high weight. The fact that such peaks are visible are related to the limited statistics of the sample.

Fig. 7.2: Shows the separation between the signal and background for the Fisher method.
Artificial neural network results

Fig. 7.3: MLP response shows the separation between the signal and background. The peaks in the background are due to events with high weights.

MLP is ranked as the best method (see Fig. 7.1). If we compare Fig. 7.2 with Fig. 7.3, it is noticeable that in the Fisher method there is not a marked separation between Higgs signal and background, while in MLP there is a tendency for the background to peak at 0 while the Higgs signal tends toward an output of 1.

Even if there is still a large overlap region below 0.5 (approximately half of the Higgs signal count in this case), there is a clear predominance of the signal related outputs to be above that value. Therefore, this method is capable of correctly classifying part of the data, even if still a large part of the signal has to be removed together with the background, or more background has to be admitted if we want to preserve the Higgs signal.
Boosted decision tree results

Fig. 7.4: Boosted decision tree response.

BDTD seem to have the second best response from the 4 qualifiers. In this case, if we compare Fig. 7.4 with 7.2 and 7.3 there is a better separation between the signal and background samples. The high peaks, is the same reason explained before.

7.1.1 Over-training check

As stated in section 6.2.3 over-training is difficult to control. In Fig. 7.5, Fig 7.6 and Fig 7.7 are shown the results of the over-training test used check whether the training has been done correctly. The Kolmogorov-Smirnov test gives a probability that quantifies the similarity between two distributions.

Fig. 7.7 shows the over-training check for boosted decision trees method. The Kolmogorov-Smirnov test presents a 4% of probability for the signal and 81 % for background. The reason for the low number in the probability for the signal can be caused by the peaks present in the plot, as explained in the subsection 7.1 due to some events with unusual high weights. Probably with a larger statistic this unusual peaks would be solved.
7. Results of optimization

Fig. 7.5: Fisher method, Kolmogorov-Smirnov test to check the over-training. The K-S probability is 29% for signal and 44% for background.

7.1.2 Sensitivity studies

In order to calculate the sensitivities for each of the four selected methods a script was written taking the outputs values from TMVA as well as the cuts. Equation 6.2.3 was used to compute this quantity.

The rectangular cut optimization output gives to each variable an upper
Boosted decision tree Fisher method, Kolmogorov-Smirnov test to check the over-training. The K-S probability is 4% for signal and 81% for background.

Table 7.1: Cuts per variable recommended by RCO

<table>
<thead>
<tr>
<th>Rectangular cut optimization</th>
<th>cuts</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>\eta_{jj}</td>
</tr>
<tr>
<td>$</td>
<td>\Delta \eta_{jj}</td>
</tr>
<tr>
<td>Mass Tagging Jets</td>
<td>[19415.52,6842457.5]</td>
</tr>
<tr>
<td>Missing Et</td>
<td>[157.39,27]</td>
</tr>
<tr>
<td>b weight</td>
<td>[-4.56,35.07]</td>
</tr>
</tbody>
</table>

The sensitivity for rectangular cut optimization is 0.6, as shown in Fig 7.8.

Fig 7.9 shows the Fisher Punzi sensitivity which yields 1.27 as the best value for sensitivity. The corresponding cut is indicated in the figure by the vertical line. Events on the right of the line are in the signal accepted region. Fig 7.10 shows the MLP Punzi sensitivity which yields 13% higher sensitivity than the Fisher at the value for the recommended cut. Finally, Fig 7.11 shows the BDTD Punzi sensitivity which is 33% better than the Fisher at its maximum value. This method gives the best sensitivity of the four used. However, there is a sharp increase in the Punzi distribution, this may be caused by the low statistic available for the study.
Fig. 7.8: Punzi sensitivity for the rectangular cut optimization method.

Fig. 7.9: Punzi sensitivity for Fisher Method. The vertical line is the cut recommended, events to the right side of the line are the allowed events.
Fig. 7.10: Punzi sensitivity for artificial neural networks method. The vertical line is the cut recommended by the MLP.

Fig. 7.11: Punzi sensitivity plot for boosted decision trees. The vertical line is the cut recommended by BDTD method, the events to the right side of the line are the allowed events.
7.1.3 Expected and optimized observables

In order to plot the remaining events after the optimization, a script with the recommended values from the multivariate techniques was made. Table 7.2 shows the sensitivity for each one of the methods.

<table>
<thead>
<tr>
<th>Method</th>
<th>RCO</th>
<th>Fisher</th>
<th>MLP</th>
<th>BDTD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Significance</td>
<td>0.6</td>
<td>1.27</td>
<td>1.44</td>
<td>1.70</td>
</tr>
</tbody>
</table>

Table 7.2: Punzi significance per method

Some characteristic plots of distributions of the observables were made. Fig. 7.12 shows the Higgs transverse mass, Fig. 7.13 invariant mass of the jets and the b-weight. In these plots the background is shown as the open histograms while the signal is presented using the filled histograms. Comparing the distributions of these observables before optimization as shown in section 5.5, one notes that the signal-background ratio has improved. However, to make the measurement significant the background must still be reduced further. For this other techniques will have to be used.

![Histogram of Transfer mass](image)

Fig. 7.12: BDTD method, Higgs transverse mass, in dashed black the Higgs signal and in white all the backgrounds. Though the shapes are different, it is possible to identify the mean of the Higgs plots around 170 GeV as expected.
Fig. 7.13: MLP method, invariant mass of the tagging jets. The Higgs signal keeps its characteristic shape.

Fig. 7.14: Fisher method, b tagging weight of the Higgs signal and background. For the Higgs signal there is not an identifiable b weight sign.
7.2 Method 2

Invariant mass and $|\eta|$ of the tagging jets, $\Delta \phi$ of the leptons.

Up to this point, the only observables considered were those related to the tagging jets. Another set of characteristic results has been obtained with mixed variables, related to leptons and jets with low correlation between the variables (Table 6.3).

Fig. 7.15 shows the four methods again but for the new configuration. It seems that MLP is giving the best signal efficiency overall. However, it is needed to do the studies per classifier.

![Fig. 7.15: TMVA response](image)

7.2.1 Methods response

From TMVA output 7.16, 7.17 and 7.18, this selection of variables looks more promising. There is a more clear separation between the signal and background in the response plots, showing the background peak around 0 and the signal one around 1.

In general the data allows a clearer separation of the signal and background. We therefore expect to see an increase in sensitivity using these variables.
Fig. 7.16: Response from MLP method, separating Higgs signal from background.

Fig. 7.17: Response BDTD Method, separating Higgs signal from background.
7. Results of optimization

Fisher response

Normalized

Signal

Background

Fig. 7.18: Fisher response for configuration between Higgs signal and background

7.2.2 Sensitivity & over-training

TMVA plots (Fig. 7.16, Fig. 7.17 and Fig. 7.18) are a good indicator of the performance of the system to correctly filter the data and separate the signal from the background. For improving sensitivity, in the other hand, it is necessary to have not only a clean data set, but also one that gives the most useful information about the studied phenomena (see Figures 7.19, 7.20 and 7.21). Therefore, as explained in section 5.5, the selection of the observables plays an important role in the system tuning process. In the Table 7.3 have been presented the different sensitivities per method.

<table>
<thead>
<tr>
<th>Punzi significance</th>
<th>RCO</th>
<th>Fisher</th>
<th>MLP</th>
<th>BDTD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.55</td>
<td>1.21</td>
<td>1.06</td>
<td>1.10</td>
</tr>
</tbody>
</table>

Table 7.3: Punzi significance per method configuration 2

This selection of variables provides data coming from the leptons, not only from the jets. Because these variables are not correlated and introduce therefore additional information, it might be expected from the response of the multivariate methods better results. And this effect is in fact reflected in the initial graphs shown in Fig. 7.15 and the methods response. However, when a large number of variables is used, it is harder to correctly tune and not over-train a multivariate analysis, and the significances obtained (see table 7.3) at the recommended values are in fact smaller than the results for
7.2 Method 2

<table>
<thead>
<tr>
<th>Rectangular cut optimization variable</th>
<th>cuts</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>\eta_{jj}</td>
</tr>
<tr>
<td>Mass Tagging Jets</td>
<td>$[25382.14,6842457.5]$</td>
</tr>
<tr>
<td>leptons $\Delta \phi$</td>
<td>$[-6.25,6.24]$</td>
</tr>
</tbody>
</table>

Table 7.4: Cuts per variable recommended by RCO method 2

Method 1 [7.2] A better treatment of the signals or a different approach in its processing might revert its behavior, like training the program with a larger number of variables than those used in method 1 and using the optimization program only to select the variables, and later for the optimization of the sensitivities. This approach was already used by [13], and is worthy of further study in this case. This shows that even if multivariate method’s response looks better it does not necessarily lead to a better result.

Fig. 7.19: Over-training and sensitivity plots for Fisher method. Left: Over-training, the K-S probability is 21% for signal and 94% for background. Right: Punzi sensitivity of 1.21$\sigma$
Fig. 7.20: Over-training and sensitivity plots for MLP method. Left: Over-training, the K-S probability is 5% for signal and 14% for background. Right: Punzi sensitivity of 1.06σ

Fig. 7.21: Over-training and sensitivity plots for BDTD method. Left: Over-training, the K-S probability is 20% for signal and 38% for background. Right: Punzi sensitivity of 1.10σ

It is recommended to study further the criteria for the detection and rejection of over-trained systems. The current criteria seems too strict and often forces the rejection of systems with a good behavior against testing data. Even if these results are too different from those of the training set, some attention should be paid to these system if they are anyway showing a good enough response against unknown samples.
A big unanswered question in particle physics is the existence or not of the Higgs boson. At the Large Hadron Collider, the ATLAS experiment has been built with the main ambition of putting together all the pieces in the puzzle and answer this question.

The purpose of the current study was to optimize the observables selection of the Vector Boson Fusion Higgs production decaying into two W’s. If the Higgs exists in the range of 170 GeV this channel is very promising, as explained. This work has shown that this channel has a clean Higgs signal characterized by two highly energetic jets in the forward part of the detector, two opposite charges leptons going in the same direction and missing $E_T$.

Knowing the characteristics of the events and using the channel properties, pre-selection cuts have been applied. The chosen cuts improved by a factor of 15 the signal-background ratio, and a clean signal can be observed.

On the remaining data, multivariate analysis tools have been used for the optimization of the sensitivity in the Higgs search. Two collections of selection observables were proposed and analyzed:

- Jets invariant mass, $|\eta_{jj}|, |\Delta \eta_{jj}|$ of the tagging jets, $E_T^{\text{miss}}$, and b-weight.

- Jet invariant mass and $|\eta|$ of the tagging jets, $\Delta \phi$ of the leptons.

Four multivariate methods have been used to optimize the sensitivity for both cases: rectangular cut method, Fisher method, artificial neural network and boosted decision tree. The results of this study indicate that a bigger number of observables to optimize gives better results as shown in Table 7.2, as long as they are correctly selected. Even if these shown results come
from particulars run cases, the general tendency observed during this works is clearly represented in them.

The most obvious finding to emerge from this study is that Boosted decision tree and artificial neural network gives the best optimization result, shown in Chapter 7 with a sensitivity of $1.70\sigma$ for Boosted decision tree and $1.44\sigma$ for artificial neural network. The non-linearity and adaptability in the methods allows to have better understanding of the signal-background relation. The tagging jets observables are sensitives observables in the search for the Higgs on the selected channel.


[6] M. Dittmar and H. Dreiner. h \( \rightarrow \) w+w- l+\( \bar{\nu}\) l-\( \nu\).


[27] Iris Rottlaender. Prospects for the search for higgs bosons with vector boson fusion processes at the lhc. 2007.


[29] www.nikhef.nl/pub/experiments/atlaswiki/index.php/AMA_in_..._14_1_0.
