Study of Kinematic Variables in Multi-Lepton Final States in search for an Exotic Higgs Boson

Axel Gallén

Supervised by Else Lytken

Thesis submitted for the degree Bachelor of Science (15 hp)

LUND
UNIVERSITY

Department of Physics

Division of Particle Physics

Spring 2021

Abstract

The Large Hadron Collider in Geneva was constructed as a discovery machine. So far, the Higgs boson has been the only discovery at the Large Hadron Collider, but physicists keep on pushing the limit in order to discover new particles. With the two major runs at the Large Hadron Collider showing some alluring surpluses of activity in multi-lepton final states, beyond the Standard Model theories are being explored in order to try to explain these excesses. One of the most prominent beyond the Standard Model theories consistent with these excesses is the 2HDM(+S) model. This theory postulates a new heavy Higgs boson H, decaying into the Standard Model Higgs boson h and a singlet scalar S. This thesis investigates the decay products of S, which are different flavour leptons of opposite sign $\ell^{\pm}\ell^{\mp}$, by studying kinematic variables in three well-defined regions at different mass points of H and S with a fixed Standard Model Higgs mass of $m_h = 125\,\text{GeV}$.

Acknowledgments

First of all, I would like to thank my supervisor Else Lytken. Your help, feedback and supervision has been invaluable this semester. The weekly meetings with you and Nathan Simpson, whom I also would like to greatly thank, have been meetings I have looked forward to every week.

I would also like to say a big thank you to Rafael Antonio Lopez. The way you explained the frameworks was illuminating, and your patience with me has been highly appreciated.

Last but not least, my sincerest gratitude goes out to my friends and family. The support and good times you have provided me with is something which have really mattered and have made a significant difference for the better. Thank you.

Contents

List of acronyms

SM	Standard Model
BSM	Beyond the Standard Model
ATLAS	A Toroidal LHC ApparatuS
CMS	Compact Muon Solenoid
2HDM(+S)	2 Higgs Doublet Model + Scalar
2HDM	2 Higgs Doublet Model
eV	electron Volts
BR	Branching Ratio
ggF	Gluon fusion
ttH	Top pair associated Higgs production
tH	Single top associated Higgs production
CERN	European Organization for Nuclear Research
LHCb	Large Hadron Collider beauty
ALICE	A Large Ion Collider Experiment
COM	Centre-Of-Mass
η	Pseudorapidity
R	Angular Separation
MET/E_T^{miss}	Missing Transverse Energy
CAF	Common Analysis Framework
MC	Monte Carlo
b-Jet	b-tagged jet
SR	Signal Region
CR	Control Region
VR	Validation Region
WP	Working Point

1 Introduction

The fundamental structure of the Universe has been a question for many centuries. There have been a lot of attempts to construct something which covers it all, but the closest physicists has gotten to a theory which describes everything rigorously has been with the theory of the Standard Model (SM). This theory was formulated during the later half of the 20th century, and ever since there have been discoveries which have established and confirmed the SM as a theory with the possibility to describe the Universe from the fundamentals. However, even though the theory has explained most of the Universe with great success, it has some flaws. As an example, the theory is not able to predict the masses of neutrinos, the unification of gravity as a force, nor the existence of dark matter.

Despite this, the success greatly outweighs the flaws mentioned above, which are the major flaws, and this leads one to continue with the SM. Furthermore, with the discovery of the Higgs boson in 2012 by the CMS [1] and ATLAS [2] detectors at the LHC, one deemed the SM complete. This means that any phenomena not explained by the SM falls into a category called Beyond the Standard Model (BSM) physics. This is the region where this thesis will take place.

The BSM theory this thesis will investigate will be a Two-Higgs Doublet Model with an extra scalar boson (2HDM(+S)), mediated by the decay $H \to Sh$. The region which will be investigated will be the di-lepton final states from the decay of the singlet scalar S and SM Higgs boson h, with opposite charge $H \to Sh \to \ell^{\pm}\ell^{\mp}$ [3, 4], where the final state leptons are products of the scalar S. This will be done by looking at kinematic variables with data recorded by the ATLAS detector at the LHC from Run 2 with a centre-of-mass energy of 13 TeV and a luminosity of $139\,\text{fb}^{-1}$.

Additionally, the thesis will use natural units where $\hbar = c = 1$ which means that energy, mass and momentum have the common unit of electron volts (eV).

2 Theory

2.1 The Standard Model

The SM is the most prominent description of the fundamental building blocks known to date. The model contains the 17 fundamental particles known to exist, as well as three of the four fundamental forces. A figure of the SM which figuratively groups the particles and forces can be seen in Figure 1, and in Sections 2.1.1 and 2.1.2, the elementary particles and the most relevant particle in this thesis, the Higgs boson, will be explained respectively.

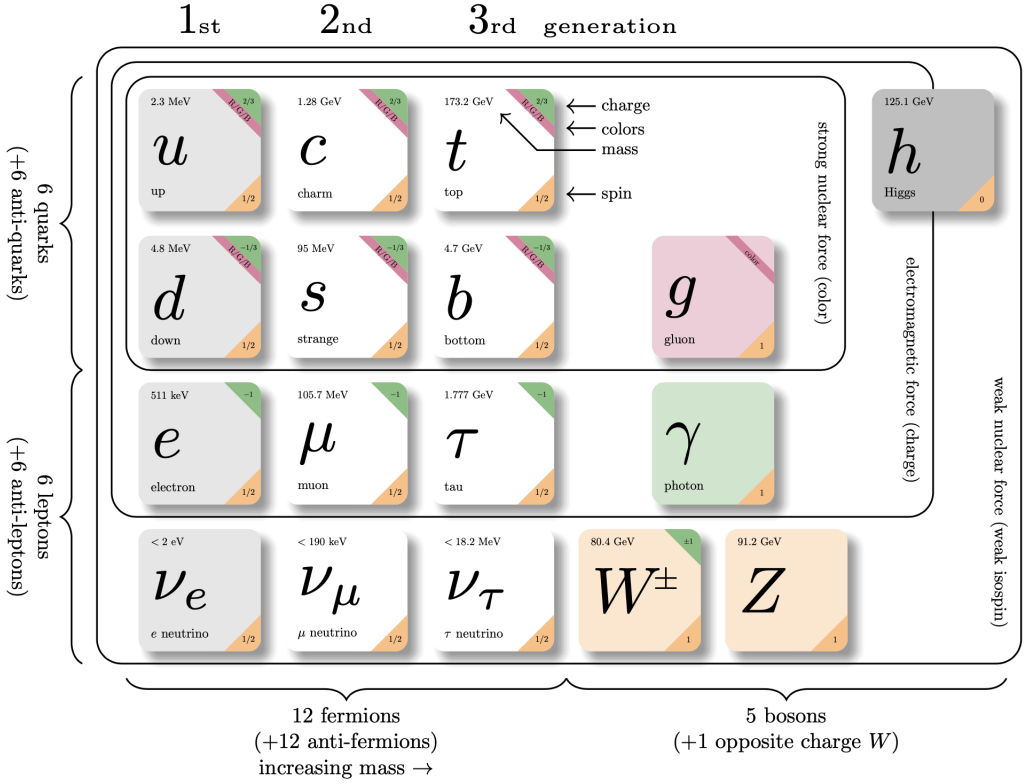


Figure 1: The Standard model which includes the elementary particles, the three forces and the particles which mediate the forces. Modified from Ref. [5].

2.1.1 The Elementary Particles

In the SM, every known elementary particle has its own place as seen in Figure 1. In the figure the different particles are divided into two sections, fermions and bosons. The fermions are particles with half integer spin $\frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \ldots$, and these particles follow so called Fermi-Dirac statistics [6]. The fermions are also divided into two families, the leptons and the quarks. In the family of leptons, the particles are the electron, muon and tau as well as their corresponding neutrino, the electron-, muon- and tau neutrino. These six particles can be grouped in generations, with the three generations being: $(e^-, \nu_{e^-})^T, (\mu^-, \nu_{\mu^-})^T, (\tau^-, \nu_{\tau^-})^T$. The leptons interact via the weak force and the electromagnetic force, while the neutrinos only interact via the weak force. The other family which makes up the fermions are the quarks. Quarks are particles which interact via the strong force, and the existing quarks, listed in Figure 1, are also grouped in three generations: $(u, d)^T, (c, s)^T, (t, b)^T$.

Physically speaking, the property which sets the leptons and the quarks apart is the property of charge. Leptons carry electrical charge and weak isospin, while the quarks carry electrical charge and colour charge. This is the factor which determines what force they interact with and, in turn, what boson interact with what fermion. This is not the only

property which differs quarks from leptons however, but there exists a hypothesis called the colour confinement hypothesis. What the hypothesis postulates is that quarks only can be found in groups such that the colour charge is neutral. This can be done in doublets, colour + anti-colour, or in triplets, colour + colour + colour. These composites are called mesons and baryons respectively, and they are bound via the strong force. In particle physics, this is a very important feature, which has a profound effect in particle collisions, namely the creation of hadronic jets.

The other particle kind in the SM are bosons. These are particles with integer spin $0, 1, 2, \ldots$, and they follow so called Bose-Einstein statistics [6]. The part which the bosons play in the SM is that they are force carriers. Looking at Figure 1, the gluon is the mediator of the strong force, the photon is the mediator of the electromagnetic force and the $W^{\pm}$ & Z bosons are the mediators of the weak force. The Higgs boson is the boson with the task of giving the other particles mass. This particle will be discussed more in Section 2.1.2.

2.1.2 The Higgs Boson

The latest addition to the SM is the Higgs boson, with it being discovered in 2012 [1, 2]. This boson's task is to give the other particles mass, which is done via the Higgs mechanism [7, 8]. The foundations of the Higgs mechanism lies in the theory of spontaneous symmetry breaking. In order to explain this, one must look at the Lagrangian which corresponds to the electroweak sector, $\mathrm{U}(1)_Y \otimes \mathrm{SU}(2)_L$, which turns out to be

$$\mathcal{L} = (D^{\mu}\Phi)^{\dagger}(D_{\mu}\Phi) - \mu^2\Phi^2 + \lambda^2\Phi^4, \tag{2.1}$$

where Φ is a doublet of complex scalar fields and μ and λ are free parameters of the SM. D_{μ} is the covariant derivative needed for the Lagrangian to be locally gauge invariant, and is defined as

$$D_{\mu} = \partial_{\mu} - ig_1\frac{Y}{2}B_{\mu} - ig_2\frac{\overline{\sigma}}{2}\cdot\overline{W}_{\mu} \tag{2.2}$$

where B^{μ} is the U(1)-hypercharge gauge field, Y is the weak hypercharge, $\overline{\sigma}$ are the Pauli matrices, $\overline{W}_{\mu}$ are the $\mathrm{SU}(2)_L$ gauge fields and $g_{1,2}$ are coupling constants for the two fields. It is also worth to notice that this Lagrangian is symmetric when $\Phi \to -\Phi$, as well as that we require $\lambda > 0$ such that the potential becomes bounded from below as $\Phi \to \infty$. The potential in its current shape has the look of a "Mexican hat", as seen in Figure 2.

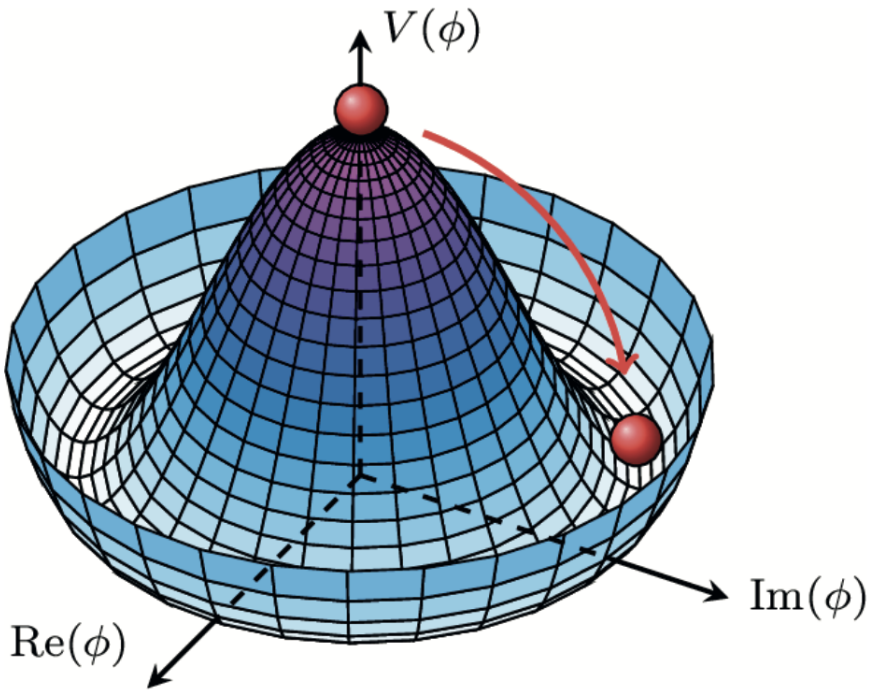


Figure 2: The Mexican hat potential representing the Higgs potential shape. Figure 2 from Ref. [9].

By appropriate choice of ground state, it can be shown that the symmetry of the Lagrangian ($\Phi \to -\Phi$) breaks. From this one obtains an important equation, namely

$$m_W = m_Z \cos\theta_W, \tag{2.3}$$

which tells one that the mass of the W and Z boson are related via the Weinberg angle, θ_W, or the electroweak mixing angle. It also says that the mass of the two (three) bosons originates from the broken symmetry, or in other words that the Higgs mechanism is the reason for the mass of the bosons.

In order to find the masses of the fermions, one must look at the Yukawa Lagrangian which originates from the doublet scalar fields (from the doublet of complex fields), and the important equation which follows from this Lagrangian is

$$m_f = \frac{y_f v}{\sqrt{2}}, \tag{2.4}$$

where m_f is the fermion mass, y_f is the Yukawa coupling strength for a fermion and v is the vacuum expectation value.

Furthermore, one important thing when talking about the Higgs is the most common decay modes of the Higgs boson, and the branching ratios for the decays. In Table 1, the most common decay modes are shown and the corresponding branching ratios.

Table 1: The most common decay channels for a Higgs boson of mass $m_h = 125\,\text{GeV}$ and the branching ratios with relative uncertainty from Ref. [10].

Decay Channel	Branching Ratio [%]	Relative Uncertainty [%]
$h \to b\bar{b}$	58.07	± 0.65
$h \to WW^*$	21.54	± 0.99
$h \to gg$	8.18	$^{+3.40}_{-3.41}$
$h \to \tau^+\tau^-$	6.26	$^{+1.17}_{-1.16}$

From this, one can clearly see that the most common decay product from the Higgs boson is the b-quark. Looking at Figure 1 however, one would not think that the bottom quark would be the most probable decay product of the Higgs, but that the top quark would be. The reason why the bottom quark is the main decay channel is due to the Higgs-fermion interaction being proportional to the fermion mass, as well as due to energy conservation. Via this argument, the top quark would be the interaction with the strongest coupling, but since it is heavier than the Higgs, conservation of energy prevents the decay from occuring. Alas, the most probable decay product is the second heaviest quark, the bottom quark.

2.2 Beyond the Standard Model

2.2.1 2HDM(+S) Model

In this thesis, a BSM theory which postulates there being one heavier Higgs boson, H, and one scalar Higgs-like particle, S, is investigated. Such a theory is called an extended Higgs sector theory, and what this particular theory postulates more precisely is two additional complex $\text{SU}(2)_L$ scalar fields, Φ_1 and Φ_2, a so-called two-Higgs doublet model (2HDM). The reason why this theory is used is because it solves one of the issues with h, namely that it is so light. This theory manages to solve this problem without fine-tuning the parameters of the SM.

For any theory postulating an SM extension with 2HDM, one can write a Lagrangian mediating said theory as [3]

$$\mathcal{L}_{\text{2HDM}} = (D^\mu \Phi_1)^\dagger (D_\mu \Phi_1) + (D^\mu \Phi_2)^\dagger (D_\mu \Phi_2) - V(\Phi_1, \Phi_2) + \mathcal{L}_{\text{int}}, \tag{2.5}$$

with $\mathcal{L}_{\text{int}}$ being a Lagrangian term containing all interaction terms, D_μ is the same derivative as in equation (2.2) and $V(\Phi_1, \Phi_2)$ is the most common scalar potential which is renormalizable.

In order to further improve the BSM theory in our case, it is advantageous to include a Higgs-like scalar particle S in our theory. The advantages with added scalar is that it is an easy way to create a connection with our BSM theory and the SM, since this scalar will couple to all SM particles. Due to its Higgs-like nature, it will also reduce the number of free parameters in the theory. This is since the BRs to the SM particles will be the same as h [3]. This makes our BSM theory a 2HDM(+S) theory, and with this new particle, two new decay processes are introduced:

$$H \to Sh \ \& \ H \to SS.$$

However, we will make an ansatz and set the branching ratio $BR(H \to Sh) = 100\%$ for simplicity reasons [11, 12]. Another advantage with this ansatz, is that it makes the process of reproducing results easier. The most usual channels where one can find this decay are shown in Figure 3, and additionally, the SM Higgs boson only decays according to the SM.

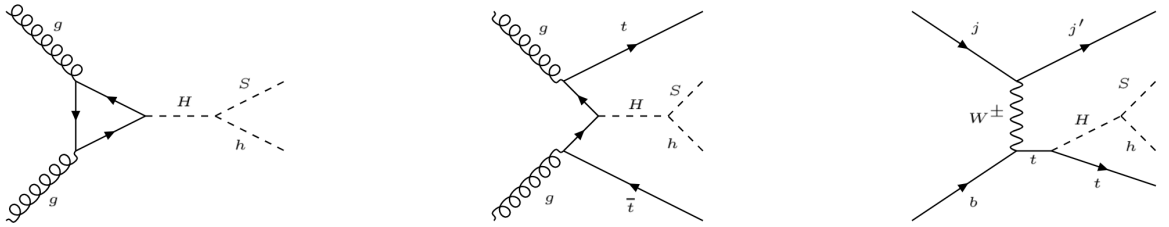


Figure 3: The three most common production modes of the heavy Higgs H, as well as its decay into S and h. Modified from Figure 1 in Ref. [11].

3 Experimental Background

3.1 The Large Hadron Collider

The LHC is the most powerfull particle accelerator and collider to date. It is considered to be the epicentre of BSM physics, and is located on the border between Switzerland and France by lake Geneva. The facility is hosted by the European Organization for Nuclear Physics (CERN), and the layout of the accelerator is a so called two-ring circular accelerator, forming a ring with a circumference of 27 km underground. This accelerator focuses on accelerating protons or heavy particles to speeds close to the speed of light, and then colliding them at four interaction points. At these four interaction points, four particle detectors with different scopes of search is located which have the possibility to record the collisions.

The four particle detectors are the LHCb (Large Hadron Collider beauty) experiment [13], the ALICE (A Large Ion Collider Experiment) experiment [14], the CMS (Compact Muon

Solenoid) experiment [15] and the ATLAS (A Toroidal LHC ApparatuS) experiment [16]. This thesis focuses on the ATLAS experiment, and it will be explained further in Section 3.2.

Furthermore, since the LHC started running, it has been upgraded a few times. This is since the energies needed to accelerate the particles to the speeds necessary is immense. The usual energy measure is referred to as centre-of-mass (COM) energy, $\sqrt{s}$, which started out at $\sqrt{s} = 7\,\text{TeV}$ when the LHC started running in 2010 (Run 1). Since then, there has been two major upgrades. One which took place in 2012 and achieved $\sqrt{s} = 8\,\text{TeV}$ [17] and marked the end of Run 1. The second upgrade marked the beginning of Run 2 which made the LHC operate at $\sqrt{s} = 13\,\text{TeV}$ between 2015 and 2018 [18].

Another relevant quantity is the luminosity, L. The luminosity is a quantity which relates the cross section, σ_{event}, to the total number of events, N_{event}. The relation between these three quantities can be written as

$$N_{\text{event}} = \sigma_{\text{event}} \int L\mathrm{d}t. \tag{3.6}$$

In other words, the integrated luminosity $\int L\mathrm{d}t$, yields the expected amount of events N_{event}, for some process with given cross section σ_{event}. In Figure 4, the total integrated luminosity of Run 2 at ATLAS is shown.

These two quantities, $\sqrt{s}$ and $\int L\mathrm{d}t$, are the two quantities which determine the performance of an accelerator. The integrated luminosity can be seen as a measure of the expected size of the data set since the quantity determines the amount of particles in the beam and the COM measures how hard the beams will collide.

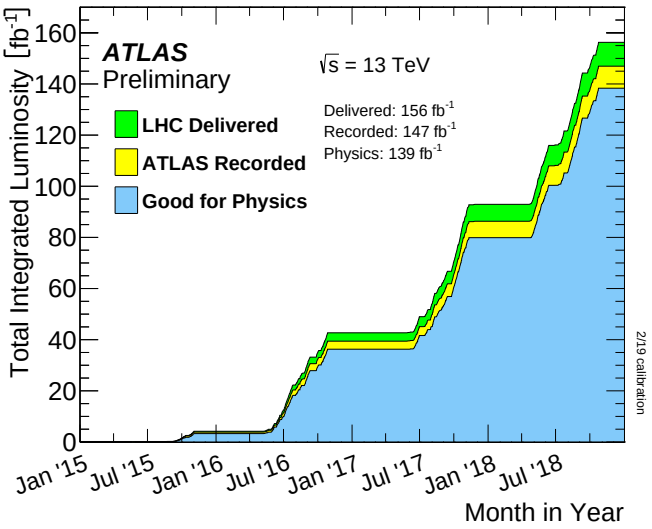


Figure 4: The total integrated luminosity, $\int L\mathrm{d}t$, measured in the ATLAS detector during the second run of the LHC at $\sqrt{s} = 13\,\text{TeV}$. Ref. [19].

3.2 ATLAS Experiment

The components of the ATLAS detector [20] are many but all necessary. The configurations of the components, or the subsystems, are arranged cylindrically around the LHC beam pipe as well as around the interaction points which is located in the geometrical centre of the detector, which is the origin in the left part of Figure 5.

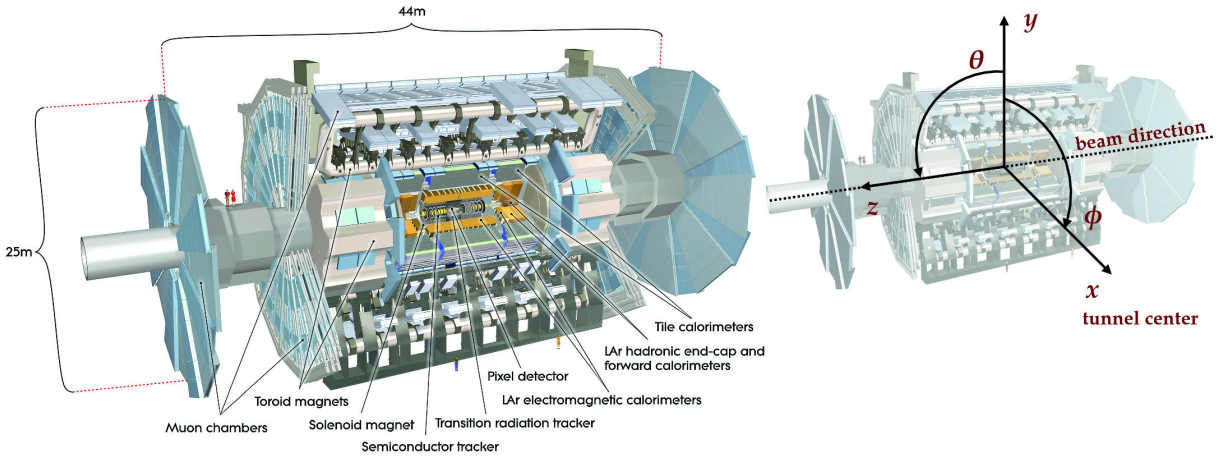


Figure 5: The figure to the left represents a cross-section view of the ATLAS detector. The figure to the right shows the coordinate system in the detector. Figure adapted from Figures 4.4 & 4.5 in Ref. [21].

Looking at the ATLAS detectors scope of search of, one sees that it is quite general compared to some of the other detectors at the LHC. One can call the ATLAS detector a general purpose detector, with some goals being to search for new heavy W- and Z-like bosons, supersymmetry and exotic Higgs bosons. This is the reason why the detector has the size and the magnitude of components which it has. Structurally, the detector has a so called *Onion Layout*, which simply means that it has many layers of detectors. Some of these detector layers are named in the left part of Figure 5 as "-chambers", "-calorimeters" and "-detector".

Furthermore, the components named "-trackers" in the figure have the job of reconstructing and identifying electrically charged particles with the help of tracking systems. This allows for measurements of particle momenta, jet tagging and identification of charged particles such as electrons and positrons.

3.3 Kinematic Variables

Such that one may properly track and measure particle quantities, a coordinate system in the detector has been defined [20]. It has been defined to be right-handed with the origin in the interaction point and the z-axis along the beam line, as shown in the right part

of Figure 5. Moreover, the coordinate system used is a cylindrical one in the transverse plane, with ϕ being the azimuthal angle measured in the x-y plane and θ being the polar angle measured in the z-y plane, i.e. from the beam axis. However, the polar angle θ is generally expressed as the pseudorapidity η, defined as

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right), \tag{3.7}$$

with the advantages being that for particles moving with small polar angles, the quantity approaches infinity, and that η becomes zero for particles moving perpendicularly to the beam line.

Combining the pseudorapidity and the azimuthal angle, another commonly used variable can be defined, namely the angular separation between particles. This quantity is defined as

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}, \tag{3.8}$$

with $\Delta\eta$ and $\Delta\phi$ being the separation in pseudorapidity and azimuthal angle between particles respectively.

Another common quantity is the missing transverse energy (MET) E_T^{miss}. This quantity is defined as the negative vectorial sum of common particles emerged from collisions [22],

$$E_T^{\text{miss}} = -\left(\sum_{i\in\mu}\mathbf{p}_T^i + \sum_{i\in e}\mathbf{p}_T^i + \sum_{i\in\gamma}\mathbf{p}_T^i + \sum_{i\in\tau}\mathbf{p}_T^i + \sum_{i\in\text{jets}}\mathbf{p}_T^i + \sum_{i\in\text{soft}}\mathbf{p}_T^i\right), \tag{3.9}$$

where $\mathbf{p}_T^i$ is the transverse momentum of each identified object. μ, e and γ are muons, electrons and photons respectively. τ are hadronically decaying τ-leptons and "soft", representing the "soft term", is reconstructed from detector signals that are not associated with any hard objects and "jets" are jets. However, there is another way to define the missing transverse energy, namely in terms of the missing energy in the x-y plane:

$$E_T^{\text{miss}} = \sqrt{(E_T^x)^2 + (E_T^y)^2}, \tag{3.10}$$

and this is the definition that will be used throughout the analysis section. The reason for the mention of equation (3.9) is that if one does not consider any soft processes, i.e. only hard processes, a quantity which estimates the event activity can be defined as H_T. This quantity is defined as the scalar sum of the transverse momenta from the reconstructed hard objects,

$$H_T = \sum p_T^\mu + \sum p_T^e + \sum p_T^\gamma + \sum p_T^\tau + \sum p_T^{\text{jets}}, \tag{3.11}$$

where

$$p_T = \sqrt{p_x^2 + p_y^2}. \tag{3.12}$$

Since we only consider processes originating from $H \to Sh \to \ell^\pm\ell^\mp$ together with jets, we can dismiss the sum over the photons, γ, and the sum over the hadronic τ in equation (3.11) to find

$$H_T = \sum p_T^\mu + \sum p_T^e + \sum p_T^{\text{jets}}. \tag{3.13}$$

4 Analysis

The analysis conducted during this thesis work will be explained in the following section together with demonstration of results. The framework used, and the frameworks it relies on, will be introduced briefly.

In order to study high energy physics properly, a lot of specialized software is required. Software which is able to separate the wanted data from the unwanted data, since particle collisions at the LHC has immense amounts of data as output.

As seen in Figure 4, the total integrated luminosity of the LHC reached a high of $156\,\text{fb}^{-1}$. $147\,\text{fb}^{-1}$ of the total luminosity was recorded by ATLAS and $139\,\text{fb}^{-1}$ of the recorded luminosity was deemed good quality data [23] at $\sqrt{s} = 13\,\text{TeV}$.

4.1 The Framework

The main framework used in this thesis is a framework built by a workgroup studying the decay process from the SM Higgs to W-bosons, $h \rightarrow WW^*$ (HWW). Firstly, this work group created the *HWWPhysicsxAODMaker* [24] [1], which has the task of processing data and simulated events prior to analysis. The data which is entered to this software is data of *AOD* and *xAOD* filetypes which are created by ATLAS' main data processor, *ATHENA* [25]. After this, the analysis itself is conducted by a framework based on the *Common Analysis Framework* (CAF) [26] [1] and *ROOT* [27] frameworks. This framework is called the *HWW software framework* [28] [1], and what makes this framework the one to use is because of its kinematic variable visualization and high energy physics analysis. What the framework does in short is that it creates a basic selection to the data. As an example, it has some demanding parameters which the particles have to have in order to consider them "good for physics". However, since not all data is relevant to this study, certain cuts and in some cases variable implementation are done manually.

The analysis however is fully done with simulated signals and background events. These are done with a handful of event generators, which will be introduced in Section 4.1.1.

4.1.1 Simulation of events: Background & Signal

In SM processes resulting in opposite-charge leptons, exceeding amounts of background is created. The most prominent background processes are top-quark events, such as $t\bar{t}, tV, t\bar{t}V$ where $V = W, Z$, diboson events like $W^{\pm}W^{\pm}/ZZ/ZW^{\pm}$ as well as triboson events like VVV. However, even though these are the most prominent ones, one must still consider potential fake background, lepton charge misidentification as well as less prominent background processes when constructing the events. The background events, as well as the

[1] Note that these are GitLab references. Unfortunately no public documentation is available.

signal samples, are constructed with Monte Carlo (MC) generators, but the different types of physics processes are constructed with different generators. As seen in Table 2, the different processes and their respective event generators are shown together with the cross section for each process.

Table 2: A summary of the background processes together with event generator origin and cross section.

Process	Event Generator	Cross section [pb]
Higgs	Powheg [29] + Pythia8 [30]	1.1020
$t\bar{t}$ & tW	Powheg + Pythia8	7297.7 & 3.997
tZ	MadGraph[31] +Pythia8	2.0736×10^{-2}
$t\bar{t}W$ & $t\bar{t}Z$	MG5_aMC@NLO[32] + Pythia8	$0.548\,22$ & 3.6864×10^{-2}
$V+\gamma$	Sherpa 2.2.8 [33]	104.86
Triboson & Diboson	Sherpa 2.2.2 [33]	7.196×10^{-3} & 12.501
W+Jets	Powheg+Pythia8	1130.6
Z+Jets	Sherpa 2.2.1 [33]	1982.2

The signal generator which generates the Heavy Higgs H and the Higgs-like scalar S are the *PYTHIA8* and *EvtGen* [34] event generators. These signals were generated as mass points which are agreeable with the 2HDM(+S) model used in the thesis. The different mass points are:

$$m_H \text{ \& } m_S = 240 \text{ \& } 170\,\text{GeV}$$
$$m_H \text{ \& } m_S = 350 \text{ \& } 240\,\text{GeV}$$
$$m_H \text{ \& } m_S = 400 \text{ \& } 240\,\text{GeV}.$$

However, in this thesis, the signal has been calibrated such that only the ggF process, see Figure 3a), is considered with a cross section $\sigma_{\text{event}} = 1\,\text{pb}$.

4.1.2 *b*-Jet tagging

Looking back at Table 1, the decay channel for the SM Higgs boson with the highest branching ratio is the $h \to b\bar{b}$ decay. Since jets are such common objects in proton-proton collisions, it is very common that Higgs bosons detected manifests through jets, which are results of hadronization of b-quarks. Since our theory wants to look at particles which have h as a decay products, jets originating from b-quarks are very compelling. Considering that this is the case, several so-called "Jet reconstruction algorithms" [35] are used which have the possibility to flavour tag jets. This means that one can figure out what quark flavour the jet originated from. One algorithm which does this is the DL1r b-tagging algorithm [36], and this is the algorithm used in this thesis.

This study interests itself in b-tagging, and this results in different operating points which the jets are evaluated at. The operating points differ from each other in such a way that the efficiency of the tagging procedure is different. In our case, the notation, with corresponding values, will be

$$\epsilon_b = (60, 70, 77, 85)\%. \tag{4.14}$$

The percentages are measures of how successful the identification of a b-tagged jet (b-jet) is as a function of their transverse momentum p_T^{jets}.

4.2 Analysis Regions

In order to achieve the best possible analysis, a good strategy is needed. This study concerns itself with a strategy which defines different analysis regions. This is done to better understand the analysis by looking at different cases in different regions, together with the possibility to optimize final states properly. Furthermore, the work done in following subsection, 4.2.1, is based on two Master theses Ref. [37, 38]. Primarily Ref. [37], which will be referred to throughout.

Additionally, the plots shown in the subsections 4.2.1, 4.2.3, 4.2.4 & 4.2.5 are all normalized to unity. This is such that one can compare the shape of the signal events to the background thoroughly.

4.2.1 Region Assessment

To begin with, the different regions investigated will be defined by so called cuts. Cuts are specific event selections which set criteria on parameters. The three different regions looked at during this thesis is the Signal Region (SR), the Control Region (CR) and the Validation Region (VR). These three regions are furthermore defined as orthogonal to each other. This means that events observed will not overlap between the regions.

In the SR, kinematic cuts to the radial separation between the final state leptons, ΔR^{ll}, the final state leptons transverse momentum, p_T^{ll} and the amount of b-jets have been applied. These cuts are applied in order to maximize the signal while minimizing the background.

Moreover, the main feature of the CR is that it only concerns high invariant mass of the final state leptons, m_{ll}, together with the same b-jet cut that the SR has. The goal of this region is to understand the background processes close to the SR better, and this is done by cutting away low m_{ll} events. This is because the high invariant mass leads to the number of signal events passing through the region being minimal, and thus negligible. However, this comes with the cost of including a lot of $t\bar{t}$-events, which can be seen as a positive if the goal is to study such events.

Finally, the VR is characterized by its incorporation of two b-jets. This is the only region containing two b-jets, together with the same invariant mass cut as in the CR. As a result

of this, we will also see a majority of $t\bar{t}$-events, together with a region where the signal is negligible. The inclusion of two b-jets will result in a more homogeneous region, and this region will also be used to justify the theoretical prediction of the background. This is done by comparing, in well-defined regions, data to simulated events.

In Table 3, the exact values which define the three regions are summarized.

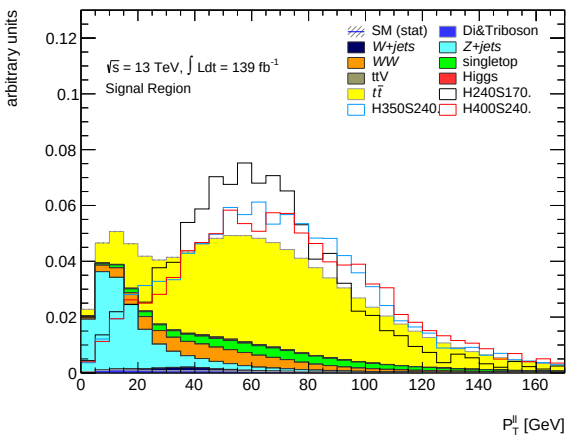
Table 3: A summary of the event selection in the three defined regions SR, CR and VR. Opted from Table 5 in Ref. [37].

Region	m_{ll} [GeV]	$n_{b-\text{jets}}$	p_T^{ll} [GeV]	ΔR^{ll}
Signal Region	-	$0, 1$	> 30	< 2.6
Control Region	> 150	$0, 1$	-	-
Validation Region	> 150	2	-	-

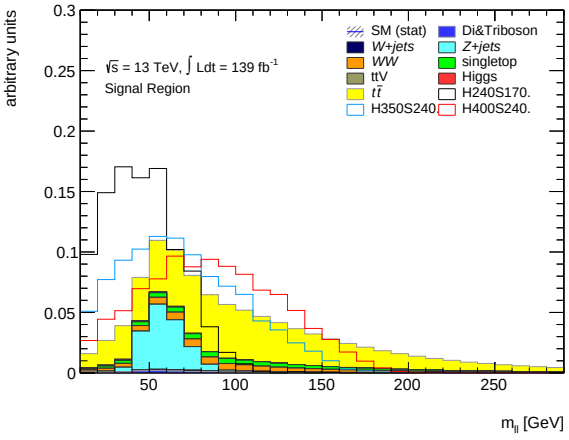
These cuts can be motivated by looking at the plots in Figure 6. It is clear that in order to maximize the signal while minimizing the background, the cuts shown in Table 3 are prominent to introduce. Additionally, the ΔR^{ll} cut can be motivated by looking at the two variables which define it (see equation (3.8)). As seen in Figure 6c,d), the majority of the background greater than 2.6 originates from the $\Delta\phi^{ll}$ variable. In this region, the signal events are also much weaker than for $\Delta\phi^{ll} < 2.6$, and the same behaviour can be seen in (c) as well.

This is not really the case for the CR and VR however. Looking at Figure 6e), the region $m_{ll} < 150\,\text{GeV}$ contains a lot more non-$t\bar{t}$ background than $t\bar{t}$ background. Since our SR lies in a region heavily dominated by $t\bar{t}$ background, we can safely dismiss this region without loosing out on new physics.

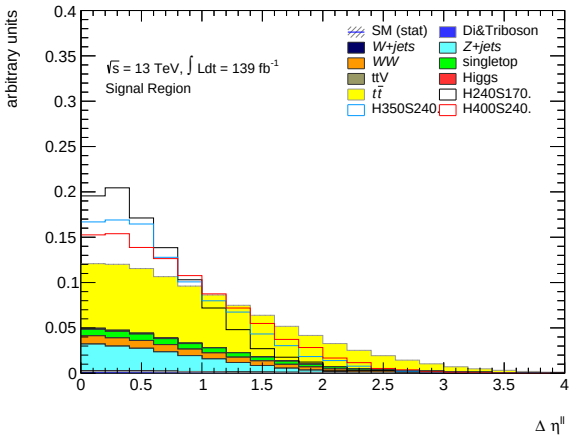
One might think that the SR should have $n_{b-\text{jets}} = 2$, since the SM Higgs most common decay is $h \to b\bar{b}$. The reason why we dismiss the two b-jets region is because $t\bar{t}$ is extra large in this region. In order to be able to extract any of information from the SR, the $n_{b-\text{jets}}$ variable have to be constrained, and a good place to make the constraint is at $n_{b-\text{jets}} = 2$.



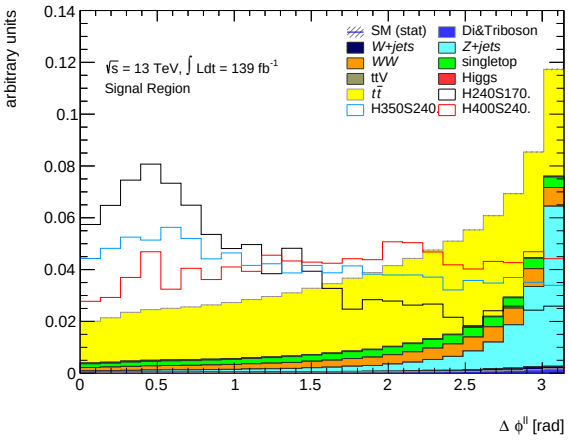
(a)



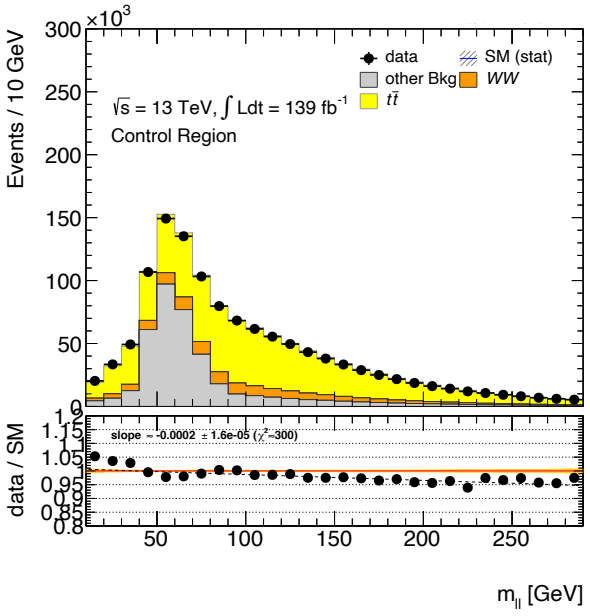
(b)



(c)



(d)



(e)

Figure 6: The transverse momentum (a), invariant mass (b), separation in pseudorapidity (c) and separation in azimuthal angle (d) of two leptons in the pre-selection. These plots are normalized to unity in order to observe the signal events against the background. The last plot, (e), shows the invariant mass of two leptons in the CR and the subplot represents the data-to-background ratio with the statistical errors displayed as the orange band.

4.2.2 Working Points

As to achieve the most optimal jet parameters, Working Points (WPs) are defined. WPs are combinations of ϵ_b (equation (4.14)) and minimum b-jet transverse momentum. The different WPs are thus combinations of b-jet tagging efficiency and minimum transverse momentum. Equation (4.15) shows the minimum transverse momentum used,

$$p_T = (20, 25, 30, 35, 40)\text{GeV}. \tag{4.15}$$

Referring back to Ref. [37], in Section 5.2, an extensive study has been done based on the Poisson significance [39],

$$\sigma_s = \sqrt{2\left[(s+b)\ln\left(\frac{b+s}{b}\right) - b\right]}, \tag{4.16}$$

where s are the signal events and b is the estimated background. The aim of the study was to determine the optimal WP as the best combination of ϵ_b and p_T in terms of σ_s. The optimal WP was determined to be

$$\text{WP}85\%p_T40.$$

This notation refers to the WP combination of $\epsilon_b = 85\%$ and $p_T = 40\,\text{GeV}$, and this is the WP which will be used during this thesis. Additionally, the different combinations of b-jets which will be looked at in the SR, CR and VR are

$$\begin{aligned} 0 &- \text{Events with 0 } b\text{-jet} \\ 1 &- \text{Events with 1 } b\text{-jet} \\ 2 &- \text{Events with 2 } b\text{-jets} \\ 0+1 &- \text{Events with either 0 or 1 } b\text{-jet}. \end{aligned}$$

4.2.3 Signal Region

As mentioned previously, the following plots are normalized to unity in order to compare the signal against the background. In Figure 7, the four variables ΔR^{ll}, H_T, p_T^{ll} and E_T^{miss}, defined as in Section 3.3, are shown in the SR. The cuts mentioned in Table 3 have been applied, and these plots also include the optimal WP together with 0+1 b-jet data.

One common theme seen throughout Figure 7, is the location of the heavier signal events (red and blue lines). They are all, more or less, located on top of the background distribution. However, one thing to notice with the lighter signals in Figure 7a,c) is that they are slightly shifted before the background peak.

Additionally, all the plots in this section are inclusive of $\mu e + e\mu$ final state leptons, where the first shown lepton is the leading lepton. Examples of plots only including μe and $e\mu$ data are shown in Appendix A.

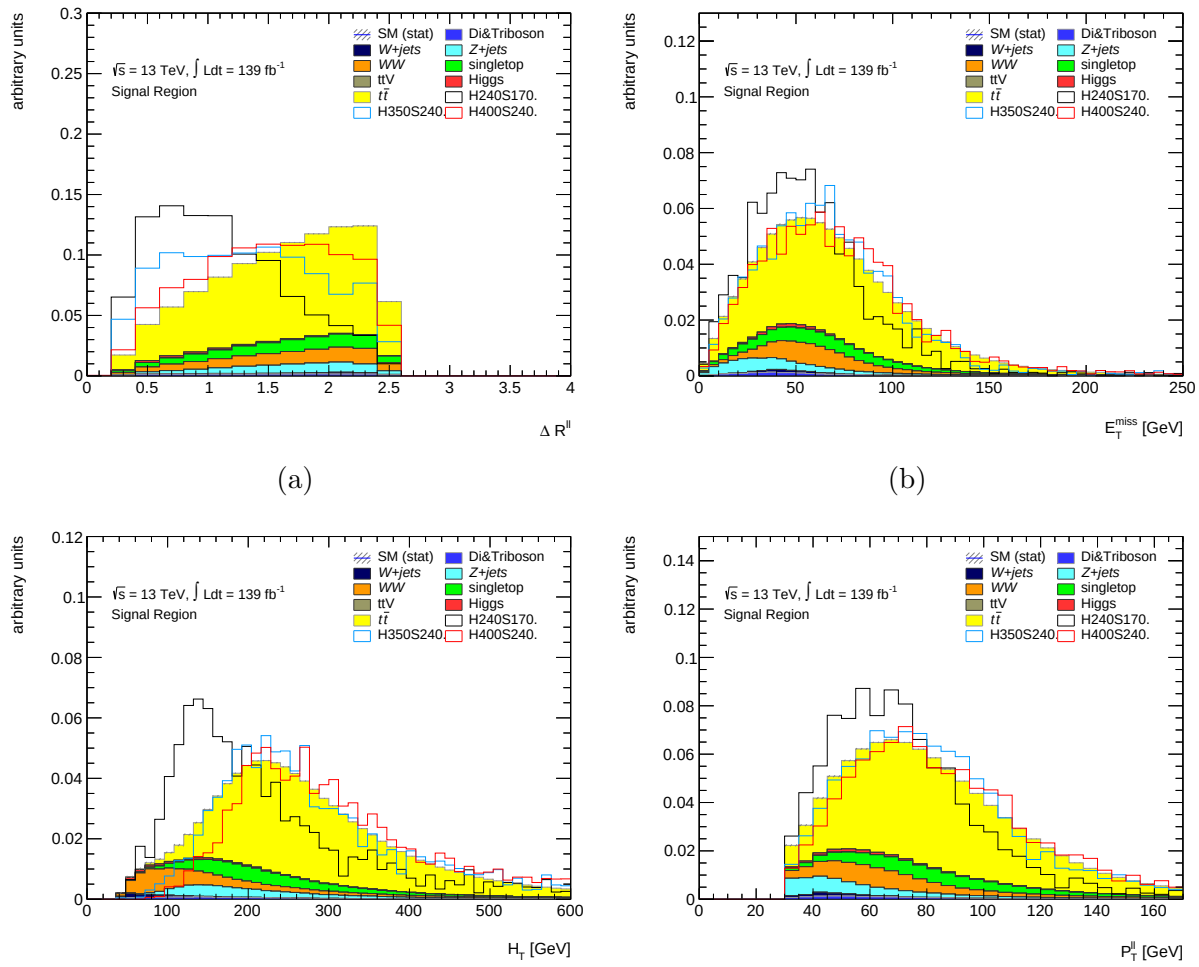


Figure 7: Four plots, angular separation (a), MET (b), hard processes (c) and transverse momentum (d) of two leptons in the SR with 0+1 b-jet data. The plots are normalized to unity in order to study the signal events against the background. The filled contours represent background and the lined contours represent signal events. The location of the signal peak compared to the background peak as well as the shape of the two are of great interest, where plot (a) and (c) are the most interesting ones.

4.2.4 Control Region

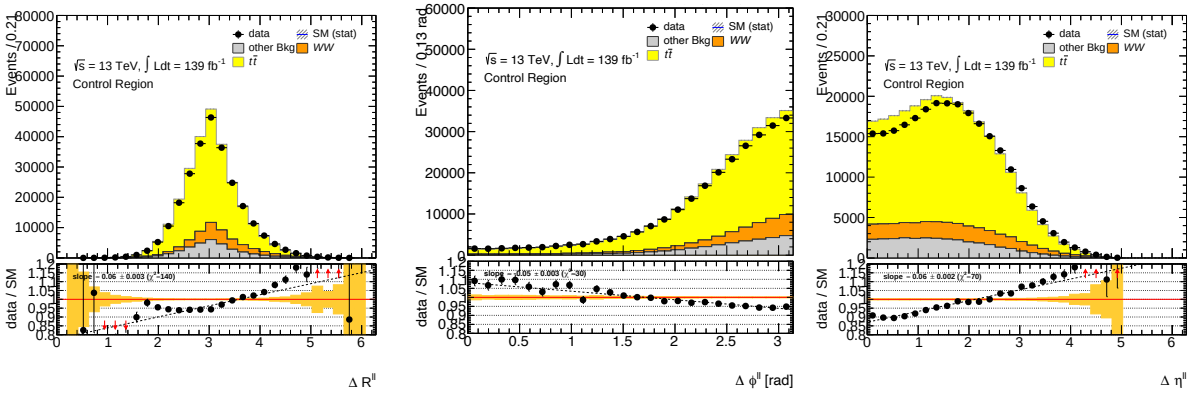
The following plots are representative of the CR, with the cuts from Table 3 applied. Furthermore, Figures 8, 9 show the kinematic variables in the CR with 0+1 b-jet region, while Figures 10 & 11 show the kinematic variables in the CR with 0 and 1 b-jet respectively. The grey filled background in the plots is named "other background". This background includes di- and triboson background (except WW), SM Higgs background and Top related background (except $t\bar{t}$).

To begin with, one can observe in Figure 8a) that ΔR^{ll} seems to be peaked around π. This is quite interesting, considering that this variable only is dependent on angles. Since the same peak can be seen in $\Delta\phi^{ll}$ as well, it is reasonable to draw the conclusion that the b-jets have travelled back-to-back, and thus also the final state leptons. Note that Figure 8b,c) will not be used further in this study, but are needed for motivational purposes.

One can also notice an increase of WW background in the 0 b-jet region, as seen in Figure 10, compared to the 1 and 0+1 b-jet regions. The explanation of this lies in that the $t\bar{t}$ background predominantly includes b-jets.

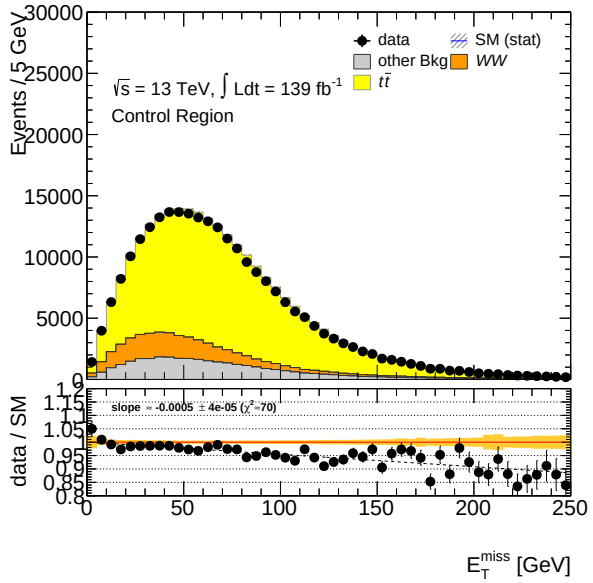
The data-to-prediction ratio also gets worse for $E_T^{\text{miss}} > 200\,\text{GeV}$. This is probably due to the lack of data and low statistics in these regions. This is also something that occurs for the $H_T \lesssim 150\,\text{GeV}$, where it especially shows in the 1 b-jet plot (see Figure 11c)).

Additionally, the data and the $t\bar{t}$ background ratio is a bit off. This is one of the improvements which can be applied to the study, which is done in Section 6 & 7 in Ref. [37]. The conclusion from that thesis, is that with the applied methods, smoothing of the data-to-background ratio was achievable, but the improvement of the final results were either quite minimal, or promising but with high uncertainty. Unfortunately, time constraints prevented further investigations in the framework of this project.

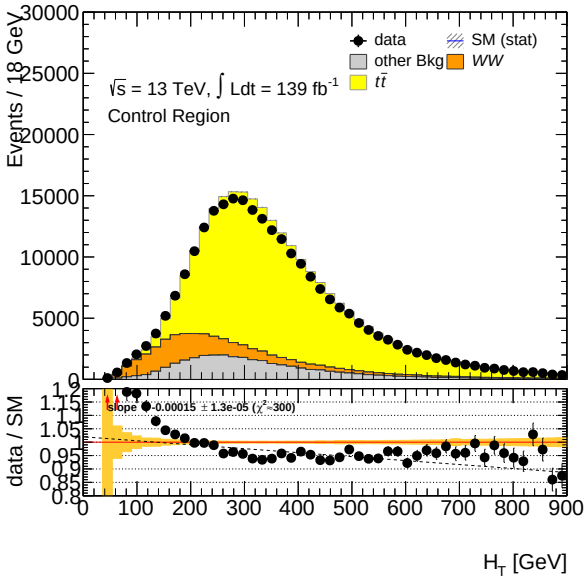


(a) (b) (c)

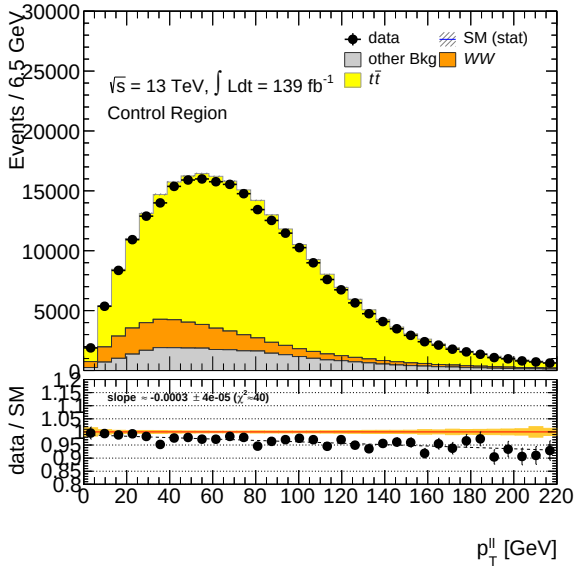
Figure 8: Three plots, angular separation (a), difference in azimuthal angle (b) and difference in pseudorapidity (c) of the two leptons in the CR with 0+1 b-jet data. Filled contours represent background and dots represent data. In the subplot, the ratio between the data and the estimated background is shown, with statistical errors displayed as the orange band. The closeness between the data and the estimated background is of importance. One can see that the closeness is varying, especially in (c) for values less than 2.



(a)



(b)



(c)

Figure 9: Three plots, MET (a), hard processes (b) and transverse momentum (c) of the two leptons in the CR with 0+1 b-jet data. The filled contours represent background and dots represent data. In the subplot, the ratio between the data and the estimated background is shown, with statistical errors displayed as the orange band. The closeness between the data and the estimated background is of importance, and one can see that the closeness is varying.

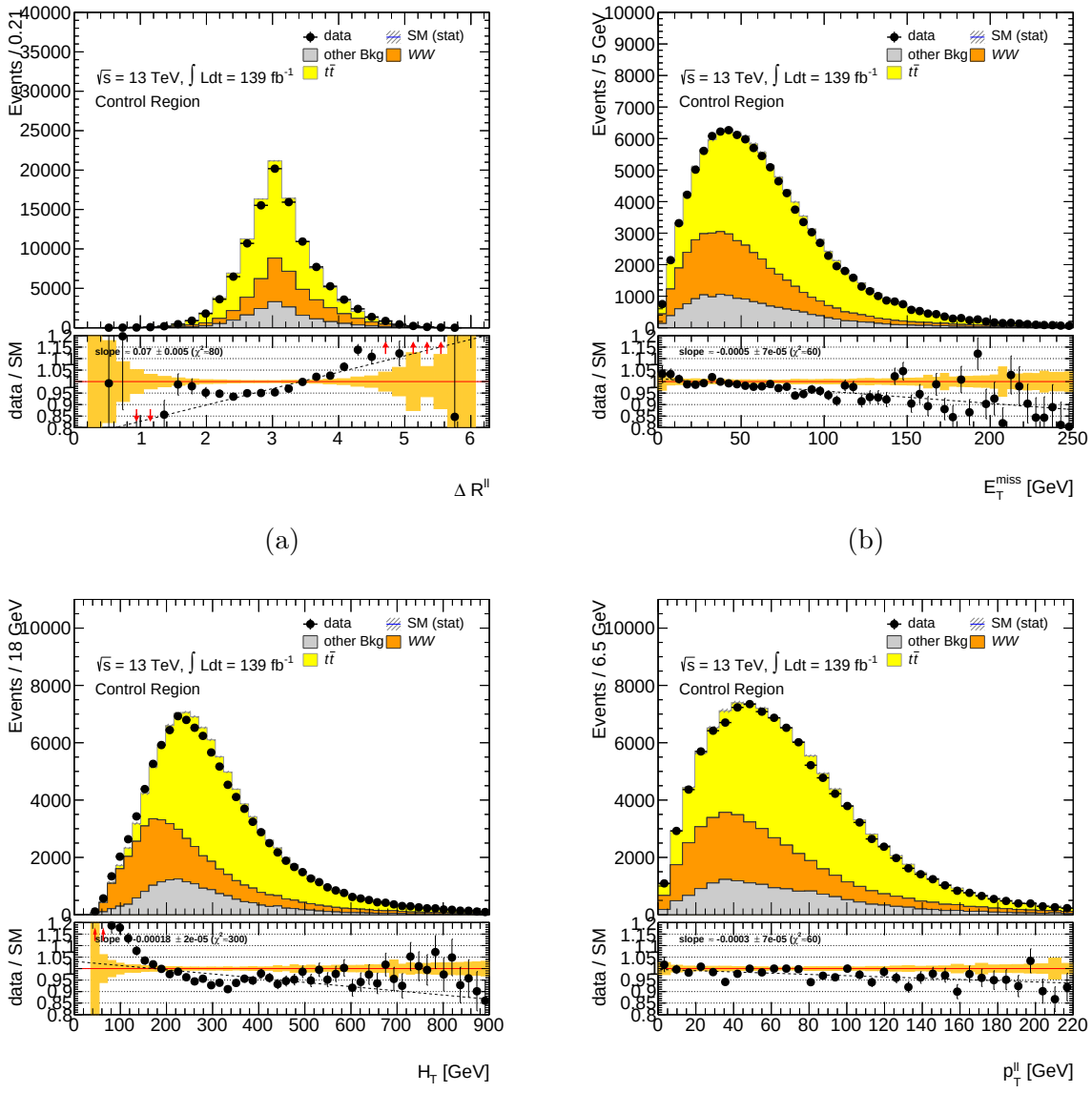


Figure 10: Four plots, angular separation (a), MET (b), hard processes (c) and transverse momentum (d) of two leptons in the CR with 0 b-jet data. The filled contours represent background and dots represent data. In the subplot, the ratio between the data and the estimated background is shown, with statistical errors displayed as the orange band. The closeness between the data and the estimated background is of importance. Plots (a), (c) and (d) especially shows this around the peaks. The large presence of WW is also interesting, motivated by the use of 0 b-jet data.

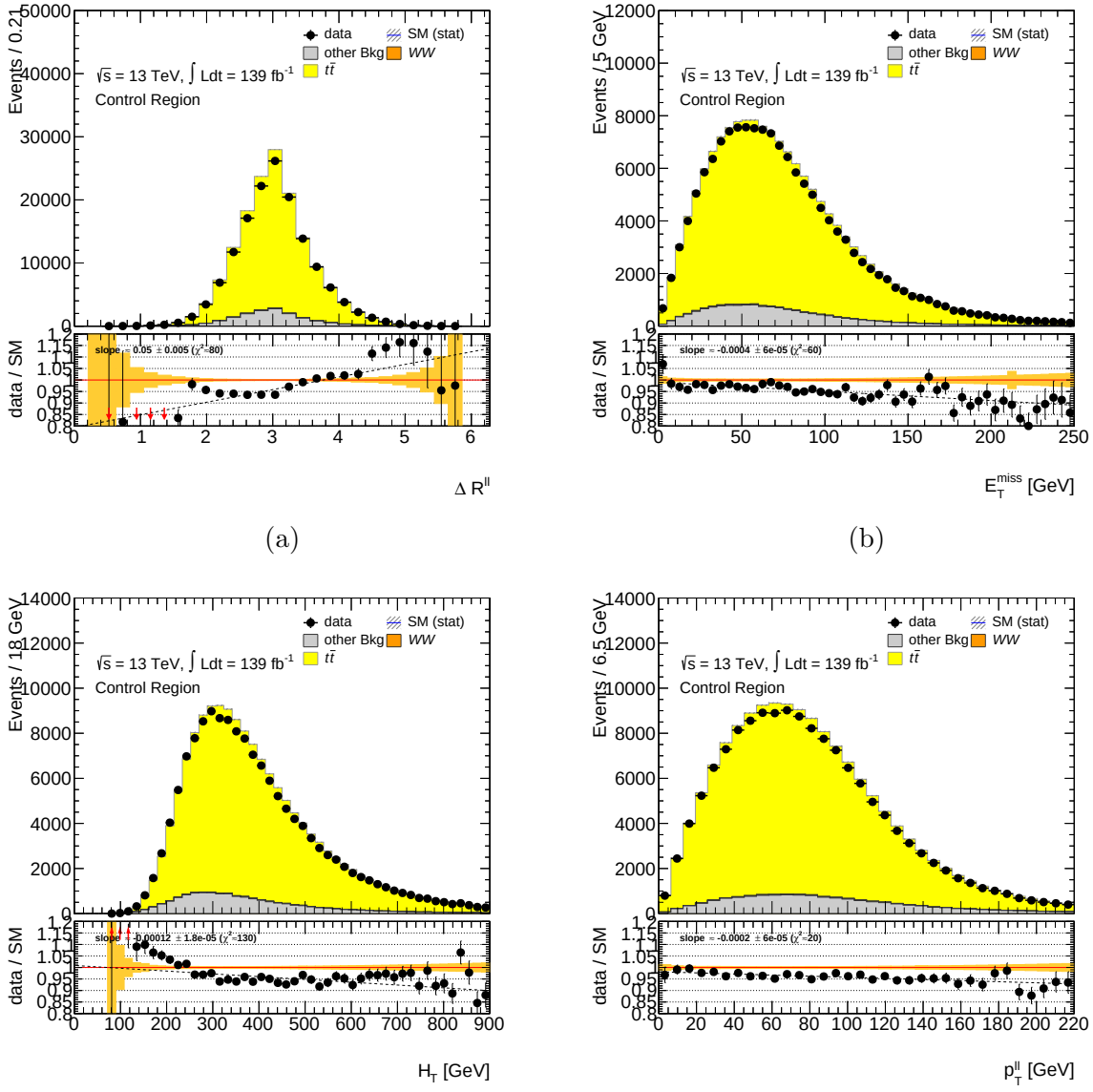


Figure 11: Four plots, angular separation (a), MET (b), hard processes (c) and transverse momentum (d) of two leptons in the CR with 1 b-jet data. The filled contours represent background and dots represent data. In the subplot, the ratio between the data and the estimated background is shown, with statistical errors displayed as the orange band. The closeness between the data and the estimated background is of importance. Plots (a), (c) and (d) especially shows this around the peaks. The great dominance of $t\bar{t}$ background is also interesting, motivated by the use of 1 b-jet data.

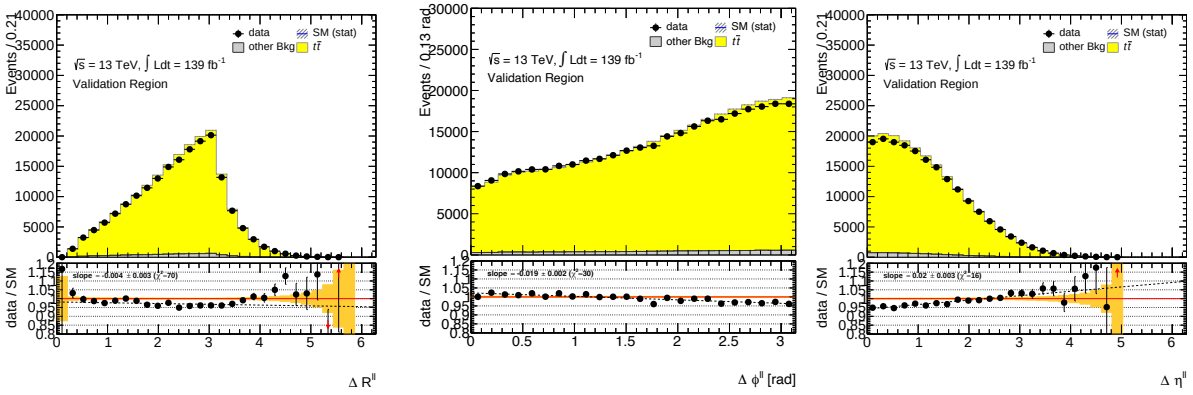
4.2.5 Validation Region

In Figures 12a) & 13, the four kinematic variables relevant to this study are shown in the VR with the cuts from Table 3 applied. Figure 12b,c) are plots necessary for motivational purposes.

As seen in the figures, this is also a region heavily dominated by $t\bar{t}$ background. One can vaguely see the same pattern here as in the CR, although not as much, with the data-to-background ratio. Comparing to the CR, ΔR^{ll} is much more behaved for low values, while the same trend for E_T^{miss} is apparent. The ΔR^{ll} variable also has a different shape than in the CR, at least for values $\lesssim \pi$, and the reason for this is clearly shown in Figure 12b). The shape is not trivially understood however as it appears linear. It is not completely understood, and definitely worth investigating further, though not investigated further due to time constraints.

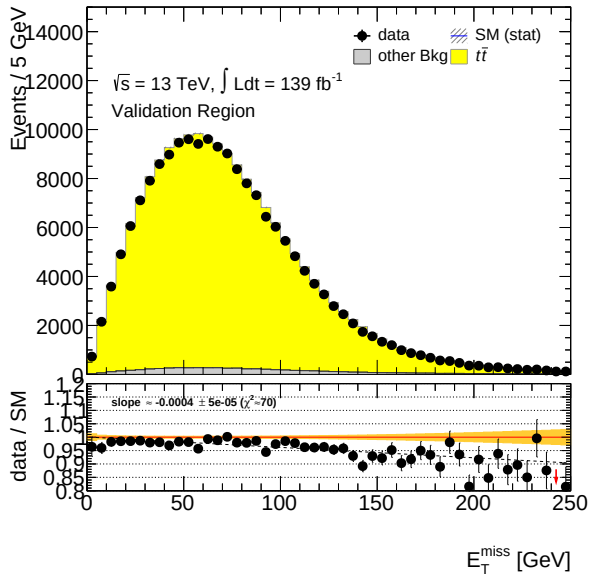
Comparing Figure 12b) to 8b), the $\Delta\phi^{ll}$ variable is much more consistent in the VR than in the CR. This is without a doubt a result of the 2 b-jets used in the VR compared to the CR.

The data-to-background ratio for p_T^{ll} is also showing the same trend as in the CR. The fact that the same trend is shown in two different regions points to it being an issue in the modelling of the simulated background.

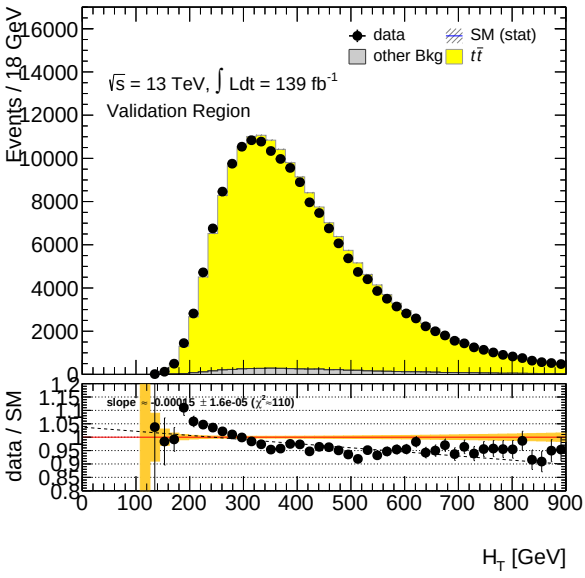


(a) (b) (c)

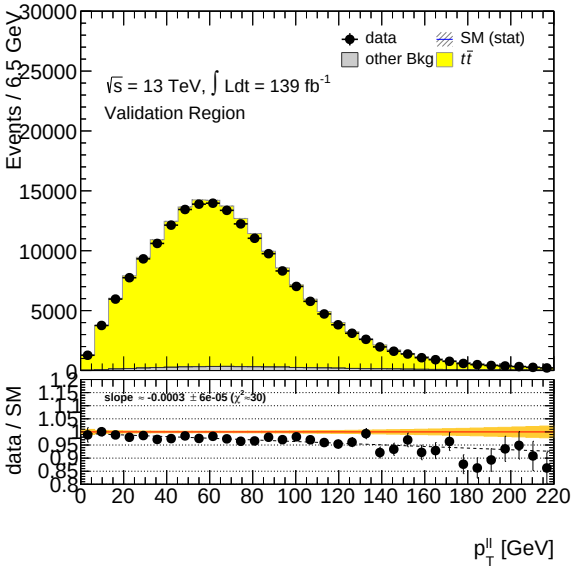
Figure 12: Three plots, angular separation (a), difference in azimuthal angle (b) and difference in pseudorapidity (c) of the two leptons in the VR with 2 b-jet data. The filled contours represent background and dots represent data. In the subplot, the ratio between the data and the estimated background is shown, with statistical errors displayed as the orange band. The closeness between the data and the estimated background is of importance. Using 2 b-jet data is the reason for the heavy $t\bar{t}$ background, and also the reason for the shape of (b).



(a)



(b)



(c)

Figure 13: Three plots, MET (a), hard processes (b) and transverse momentum (c) of the two leptons in the VR with 2 *b*-jet data. The filled contours represent background and dots represent data. In the subplot, the ratio between the data and the estimated background is shown, with statistical errors displayed as the orange band. The closeness between the data and the estimated background is of importance. The greatest deviations in closeness are seen in (a) and (b) in the values around the peak in each plot.

5 Conclusion and Summary

The aim of this thesis was to investigate a 2HDM(+S) BSM model. This was done by studying kinematic variables in multi-lepton final states from pp-collision data, with $\sqrt{s} = 13\,\text{TeV}$ and $\int L\text{d}t = 139\,\text{fb}^{-1}$, at the LHC with the ATLAS detector. The decay channel of this BSM model in interest is $H \to Sh \to \ell^{\pm}\ell^{\mp}$, where the final state leptons of opposite charge are decay products of the heavy Higgs-like scalar S. The particle one hopes to find is the BSM Higgs boson H and the scalar particle S, but due to the nature of the bosons and their final states, which are similar to wide range of SM processes, it is a particularly difficult region to investigate. Especially due to the colossal $t\bar{t}$ background prevailing in this region. This thesis therefore relies heavily on well optimized selections and tools for analysis to make the process as smooth as possible.

Furthermore, this thesis investigates the postulated new bosons at three different mass points in three different well-defined regions, with a total of four kinematic variables. The three regions are defined by a series of pre-selection cuts on the kinematic variables and on the amount of b-jets in the regions.

Based on the analysis in the three different regions, one conclusion one was able to draw was that out of the three mass points, the lightest one was the one which deviated from the background the most, especially clear in the H_T and ΔR^{ll} plots. Another conclusion was that the $t\bar{t}$ background is not well-described. This was a known issue however, which has been looked into (see Ref. [37]), but due to time limitations, it was not possible to implement this into this thesis.

5.1 Outlook

Based on this study, possibilities to further improve and delve into this study exists. As an example, one could look into same leptonic final states with different charge as well as same leptonic final states with same charge. One could also look at a larger selection of kinematic variables in hope of finding variables where the signal and the background peak at different values, or even by defining the b-jet working points at higher transverse momentum to get a narrower look, even though it very likely will come with an increase of background.

A Appendix

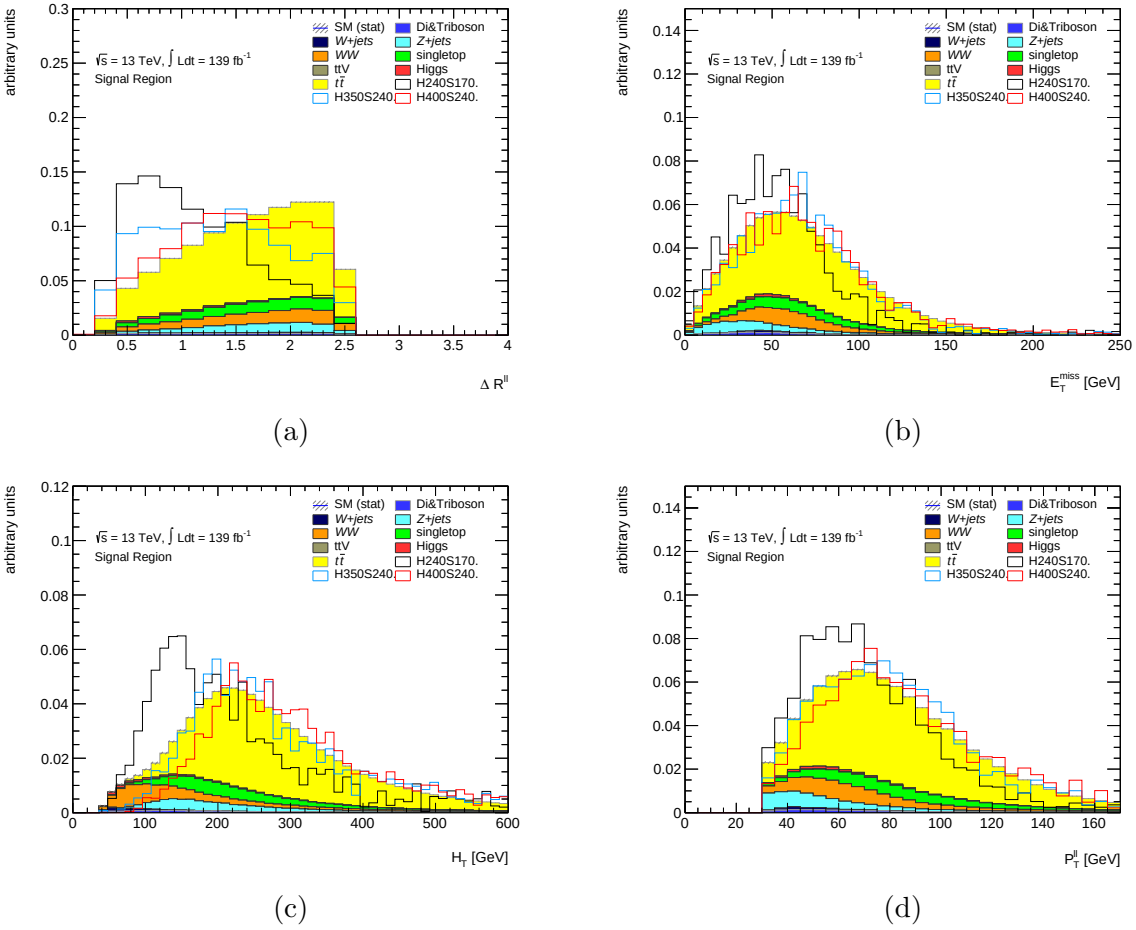


Figure 14: Four plots, angular separation (a), MET (b), hard processes (c) and transverse momentum (d) of two leptons in the SR with 0+1 b-jet data. The plots are normalized to unity, and the filled contours represent background and the lined contours represent signal events. These plots only include leading electron and sub-leading muon data, $e\mu$.

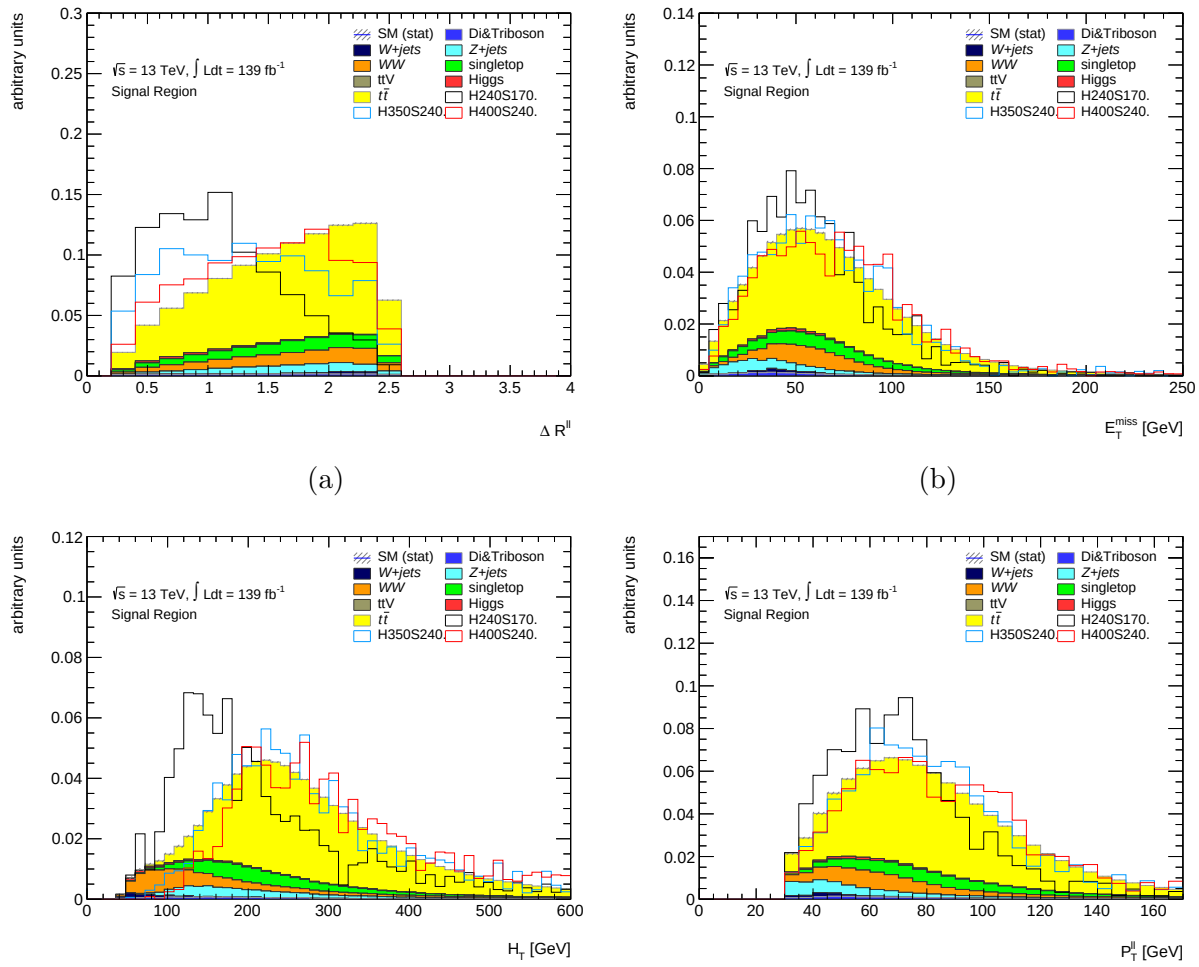


Figure 15: Four plots, angular separation (a), MET (b), hard processes (c) and transverse momentum (d) of two leptons in the SR with 0+1 b-jet data. The plots are normalized to unity, and the filled contours represent background and the lined contours represent signal events. These plots only include leading muon and sub-leading electron data, μe.

References

[1]CMS collaboration, "Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc", Physics Letters B **716**, 30–61 (2012).

[2]ATLAS collaboration, "Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc", Physics Letters B **716**, 1–29 (2012).

[3]S. von Buddenbrock, N. Chakrabarty, A. S. Cornell, D. Kar, M. Kumar, T. Mandal, B. Mellado, B. Mukhopadhyaya, R. G. Reed, and X. Ruan, "Phenomenological signatures of additional scalar bosons at the lhc", The European Physical Journal C **76**, 10.1140/epjc/s10052-016-4435-8 (2016).

[4]S. von Buddenbrock, A. S. Cornell, E. D. R. Iarilala, M. Kumar, B. Mellado, X. Ruan, and E. M. Shrif, "Constraints on a 2hdm with a singlet scalar and implications in the search for heavy bosons at the lhc", Journal of Physics G: Nuclear and Particle Physics **46**, 115001 (2019).

[5]C. Burgard, *Example: standard model physics*, (2016) https://texample.net/tikz/examples/model-physics/ (visited on 03/22/2021).

[6]Y. Kim, "Spin and statistics of elementary particles", in *Mathematical foundations of quantum theory*, edited by A. Marlow (Academic Press, 1978), pp. 347–349.

[7]P. W. Higgs, "Broken symmetries and the masses of gauge bosons", Phys. Rev. Lett. **13**, 508–509 (1964).

[8]F. Englert and R. Brout, "Broken symmetry and the mass of gauge vector mesons", Phys. Rev. Lett. **13**, 321–323 (1964).

[9]K. Mankinen, "Search for doubly charged higgs bosons in multi-lepton final states with the atlas detector", eng, PhD thesis (Lund University, Jan. 2020).

[10]D. de Florian, C. Grojean, F. Maltoni, C. Mariotti, et al., *Handbook of lhc higgs cross sections: 4. deciphering the nature of the higgs sector*, 2017.

[11]S. von Buddenbrock, A. S. Cornell, Y. Fang, A. F. Mohammed, M. Kumar, B. Mellado, and K. G. Tomiwa, "The emergence of multi-lepton anomalies at the lhc and their compatibility with new physics at the ew scale", Journal of High Energy Physics **2019**, 10.1007/jhep10(2019)157 (2019).

[12]S. von Buddenbrock, A. S. Cornell, A. Fadol, M. Kumar, B. Mellado, and X. Ruan, *Multi-lepton signatures of additional scalar bosons beyond the standard model at the lhc*, 2018.

[13]A. A. Alves, L. M. Andrade, F Barbosa-Ademarlaudo, I Bediaga, et al., "The LHCb Detector at the LHC", JINST **3**, Also published by CERN Geneva in 2010, S08005 (2008).

[14]K Aamodt, A Abrahantes Quintana, R Achenbach, S Acounis, D Adamova, C Adler, M Aggarwal, F Agnese, G Aglieri Rinella, et al., "The ALICE experiment at the CERN LHC. A Large Ion Collider Experiment", JINST **3**, Also published by CERN Geneva in 2010, S08002. 259 p (2008).

[15]S Chatrchyan, G Hmayakyan, V Khachatryan, A. M. Sirunyan, R Adolphi, G Anagnostou, R Brauer, W Braunschweig, H Esser, L Feld, W Karpinski, A Khomich, et al., "The CMS experiment at the CERN LHC. The Compact Muon Solenoid experiment", JINST **3**, Also published by CERN Geneva in 2010, S08004. 361 p (2008).

[16]G Aad, S Bentvelsen, G. J. Bobbink, K Bos, H Boterenbrood, G Brouwer, E. J. Buis, J. J. F. Buskop, A. P. Colijn, R Dankers, C Daum, R de Boer, P de Jong, P Ennes, et al., "The ATLAS Experiment at the CERN Large Hadron Collider", JINST **3**, Also published by CERN Geneva in 2010, S08003. 437 p (2008).

[17]R Alemany-Fernandez, E Bravin, L Drosdal, A Gorzawski, V Kain, M Lamont, A Macpherson, G Papotti, M Pojer, L Ponce, S Redaelli, G Roy, M Solfaroli Camillocci, W Venturini, and J Wenninger, *Operation and Configuration of the LHC in Run 1*, (2013) `https://cds.cern.ch/record/1631030` (visited on 04/10/2021).

[18]J. Wenninger, *Operation and Configuration of the LHC in Run 2*, (2019) `https://cds.cern.ch/record/2668326` (visited on 04/10/2021).

[19]ATLAS collaboration, *ATLAS experiment - Public results*, `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2` (visited on 04/15/2021).

[20]The ATLAS Collaboration, "The ATLAS experiment at the CERN large hadron collider", Journal of Instrumentation **3**, S08003–S08003 (2008).

[21]A. Tarek Abouelfadl Mohamed, "The lhc and the atlas experiment", in *Measurement of higgs boson production cross sections in the diphoton channel: with the full atlas run-2 data and constraints on anomalous higgs boson interactions* (Springer International Publishing, Cham, 2020), pp. 61–101.

[22]*Object-based missing transverse momentum significance in the ATLAS detector*, tech. rep. ATLAS-CONF-2018-038 (CERN, Geneva, 2018).

[23]ATLAS collaboration, *ATLAS experiment - Public results*, `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/RunStatsPublicResults2010#Full_Run_2_period_2015_2018` (visited on 04/16/2021).

[24]HWW, *HWWPhysicsxAODMaker*, `https://gitlab.cern.ch/atlas-physics/higgs/hww/HWWPhysicsxAODMaker` (visited on 04/15/2021).

[25]*ATLAS Computing: technical design report*, Technical design report. ATLAS (CERN, Geneva, 2005).

[26]CAF, *Common Analysis Framework*, `https://gitlab.cern.ch/atlas-physics/higgs/hww/HWWAnalysisCode` (visited on 04/15/2021).

[27]I. Antcheva, M. Ballintijn, B. Bellenot, M. Biskup, R. Brun, N. Buncic, P. Canal, D. Casadei, O. Couet, V. Fine, and et al., "Root — a c++ framework for petabyte data storage, statistical analysis and visualization", Computer Physics Communications **180**, 2499–2512 (2009).

[28]HWW, *HWWAnalysisCode*, `https://gitlab.cern.ch/atlas-physics/higgs/hww/HWWAnalysisCode` (visited on 04/15/2021).

[29]P. Nason and C. Oleari, "Nlo higgs boson production via vector-boson fusion matched with shower in powheg", Journal of High Energy Physics **2010**, `10.1007/jhep02(2010)037` (2010).

[30]T. Sjöstrand, S. Ask, J. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. Rasmussen, and P. Skands, "An introduction to pythia 8.2", English, Computer Physics Communications **191**, 159–177 (2015).

[31]J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, "Madgraph 5: going beyond", Journal of High Energy Physics **2011**, `10.1007/jhep06(2011)128` (2011).

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

[33]E. Bothmann, G. S. Chahal, S. Höche, J. Krause, F. Krauss, S. Kuttimalai, S. Liebschner, D. Napoletano, M. Schönherr, H. Schulz, S. Schumann, and F. Siegert, *Event generation with sherpa 2.2*, 2019.

[34]D. J. Lange, "The EvtGen particle decay simulation package", Nucl. Instrum. Meth. A **462**, edited by S. Erhan, P. Schlein, and Y. Rozen, 152–155 (2001).

[35]R. Atkin, "Review of jet reconstruction algorithms", Journal of Physics: Conference Series **645**, 012008 (2015).

[36]G. Aad, B. Abbott, D. C. Abbott, A. A. Abud, K. Abeling, D. K. Abhayasinghe, S. H. Abidi, O. S. AbouZeid, N. L. Abraham, and et al., "Atlas b-jet identification performance and efficiency measurement with ttbar events in pp collisions at sqrt(s)=13 tev", The European Physical Journal C **79**, `10.1140/epjc/s10052-019-7450-8` (2019).

[37]R. A. Lopez, "Search for higgs-like bosons through optimization of top-quark modelling", eng, Student Paper (2021).

[38]M. Burman Ingeberg, *Optimization of event selection and fake background estimation in a search for heavy scalars in dilepton final states with the atlas detector*, eng, Student Paper, 2021.

[39]G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", The European Physical Journal C **71**, `10.1140/epjc/s10052-011-1554-0` (2011).