GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN IN PUBLICA COMMODA SEIT 1737

Fakultät für Physik

Bachelor's Thesis

Untersuchung der Modellierung des W+jets Untergrundes im $HH/SH \to b\bar{b}VV^*$ Zerfallskanal mit einem Lepton im Endzustand mit dem ATLAS Detektor

Study of the modelling of the W+jets background in the $HH/SH \to b\bar{b}VV^*$ decay channel with one lepton in the final state using the ATLAS detector

prepared by

Zobeyer Ghafoor

from Göttingen

at the II. Physikalischen Institut

Thesis number:	II.Physik-UniGö-BSc-2023/01
Thesis period:	17th October 2022 until 6th February 2023
First referee:	Prof. Dr. Stan Lai
Second referee:	Prof. Dr. Steffen Schumann

Abstract

In dieser Arbeit wird der simulierte W+jets Untergrund der Suche nach $X \to SH/HH \to b\bar{b}VV^*$ mit einem Lepton im Endzustand untersucht. Dazu wird dieser mit Daten in der Validierungs- und W+jets Kontrollregion verglichen. Es wurde beobachtet, dass für verschiedene Variablen der simulierte W+jets Untergrund große Abweichungen zu den Daten besitzt. Um diese Abweichungen zu minimieren und die Modellierung zu verbessern, werden Gewichte auf die W+jets Untergrund Events angewandt. Diese Gewichte stammen von der W+jets Verteilung von anderen Variablen in der W+jets Kontrollregion. Die wichtigste Variable in dieser Arbeit ist die Masse $m_{\text{vis+met}}^{SH/HH}$, die die größte Abweichung des W+jets Hintergrunds von den Daten aufweist. Die Modellierung verbessert sich durch die Gewichte, welche aus der Verteilung des transversalen Impulses $p_T^{H\to b\bar{b}}$ des $H \to b\bar{b}$ Kandidaten abgeleitet werden.

Stichwörter: Higgs Boson, Standardmodell, resonante Higgs-Boson-Paarproduktion, W+jets Untergrund, Sherpa

Abstract

This thesis studies the simulated W+jets background in the search for $X \to SH/HH \to b\bar{b}VV^*$ with one lepton in the final state. It is compared to data in the validation region and W+jets control region. A disagreement between the W+jets background and data is observed for various variables. To improve the modelling of the W+jets background, weights are applied to the W+jets background events. These weights are obtained from distributions of different variables in the W+jets control region. In this thesis, the most important variable is the visible+met mass $m_{\text{vis+met}}^{SH/HH}$. It exhibits the biggest disagreement between W+jets background and data. The modelling improves when applying weights obtained from the distribution of the transverse momentum $p_T^{H\to b\bar{b}}$ of the $H \to b\bar{b}$ candidate.

Keywords: Higgs boson, Standard Model, resonant Higgs boson pair production, W+jets background, Sherpa

Contents

1. Introduction

For centuries, physicists have strived to discover the laws of nature. With the Standard Model (SM) of particle physics [1–4], they have managed to provide a solid explanation for the mechanisms of the elementary particles, the building blocks of the universe. The SM describes all the elementary particles known to date and summarizes their properties and interactions into a unified theory.

The latest achievement of the SM was the discovery of the Higgs boson in 2012 at the Large Hadron Collider (LHC) [5] and its two biggest detectors ATLAS [6] and CMS [7] at CERN, after it was theoretically predicted by Higgs [8] and Englert and Brout [9] in the 1960's. This provided experimental evidence to the corresponding Higgs field and masses of elementary particles which result from the particles coupling to the Higgs field and is a key prediction of the SM.

Since then many tests have been performed on the Higgs boson, confirming its SM-like nature. The final test, verifying the Higgs potential, remains open. This can be done by measuring the Higgs boson pair production cross section.

The SM, while correctly making many precise predictions, also has some shortcomings. One of the open questions is the matter-antimatter asymmetry in the universe [10] that has been observed. After the Big-Bang, an equal amount of matter and antimatter should have been created but almost everything that is observable to this day consist of matter. The SM does not give an explanation to this imbalance and therefore, it is necessary to test advanced theories beyond the SM which could give an answer to this fundamental question.

In principle, a pair of Higgs bosons can also be produced from the decay of a massive boson that is predicted by theories beyond the Standard Model. To conduct searches for such particles, it is crucial to understand the background processes and their modelling to distinguish them from signal events. In this thesis, the focus is set on the W+jets background as one major background in the boosted $X \to SH/HH \to b\bar{b}WW^*$ search with 1 lepton in the final state.

The thesis is structured as follows: Chapter 2 outlines the theoretical foundation of particle physics including the Higgs boson pair production. In Chapter 3, the Large

Hadron Collider (LHC) and the ATLAS detector with its components are described in detail. This is followed by Chapter 4, where the analysis steps for the boosted decay channel $X \to SH/HH \to b\bar{b}WW^*$ with 1 lepton in the final state are explained. Chapter 5 discusses into the background modelling in the W+jets Control Region (CR) and how to improve it, while Chapter 6 gives an insight into the background modelling in the Validation Region (VR) of the analysis. The thesis is closed by a conclusion and outlook in Chapter 7.

2. The Standard Model of Particle Physics

This chapter gives a short introduction to the theory of the Standard Model of particle physics (SM). It describes the elementary particles and their fundamental interactions with each other. The four fundamental forces in physics are the strong, weak, electromagnetic and gravitational force, but the latter is not described by the SM due to the weakness of this force between the elementary particles.

2.1. Elementary particles and their interactions

The SM classifies the elementary particles into two parts: the fermions as matter particles and the bosons as mediators of the fundamental interactions. Fermions have a spin of 1/2 (in units of the reduced Planck constant $\hbar$) and bosons are spin-1 particles. The Higgs boson is a scalar (spin-0) boson.

The fermions consist of six leptons that are divided into three families which can be seen in Table 2.1. Each lepton family consists of a charged lepton and its neutrino partner. Six quarks make up the other half of the fermions that can also be seen in Table 2.1. They are also divided into three generations. The electron e^-, muon μ^- and tau τ^- lepton all have an electric charge of $Q = -1$ (in units of the electric charge e), while the neutrinos ν_e, ν_μ and ν_τ do not carry electric charge. The up-, charm- and top-quark have an electric charge of $Q = +2/3$, the other three quarks possess an electric charge of $Q = -1/3$. Every particle has an antiparticle with the same properties but with an opposite electric charge.

As mentioned before, the gauge bosons are the mediators of interactions. Particles that carry the charge of the corresponding force can interact via that force. The three forces that are described by the SM have their own mediator(s) resulting in five gauge bosons: the photon γ, the gluon g, the W^+ and W^- boson and the Z boson. The sixth boson in the SM, the Higgs boson H, is not a gauge boson.

The massless and electrically neutral photon as the mediator of the electromagnetic

	Fermions			Bosons
Quarks	up (u)	charm (c)	top (t)	γ
	down (d)	strange (s)	bottom (b)	g
Leptons	ν_e	ν_μ	ν_τ	$W^\pm$
	e^-	μ^-	τ^-	Z^0
				H

Table 2.1.: The Standard Model of particle physics [11].

force couples to fermions with electric charge. In contrast to the photon, the massless gluon is the mediator of the strong force and itself carries the colour charge of the strong force. Only quarks carry the colour charge and therefore interact strongly. In the SM, there are three colour charges: red (r), blue (b) and green (g). Compounds of the quarks form colourless hadrons which can be divided into baryons (with a spin of $\pm 1/2$ or $\pm 3/2$) and mesons (with a spin of ± 1 or 0). A baryon is formed out of three quarks and a meson consists of a quark and an antiquark. Additionally, only colourless particles can propagate freely, which is known as colour confinement. The $W^\pm$ and Z bosons are the mediators of the weak interaction and all fermions interact through the weak force because they possess the weak charge. The $W^\pm$ bosons only couple to particles with left handed chirality and to antiparticles with right handed chirality.

In quantum field theory (QFT) [12, 13], the SM is a gauge invariant quantum field based on the symmetry group $SU(3) \otimes SU(2)_L \otimes U(1)_Y$[1]. This means that the Lagrangian $\mathcal{L}$ is invariant under transformations contained in $SU(3) \otimes SU(2)_L \otimes U(1)_Y$ group.

2.2. Higgs mechanism

The electromagnetic and weak interactions have been brought together via electroweak unification [2–4] that is based on the group $S(2)_L \otimes U(1)_Y$. The simplest version of the electroweak theory predicts massless $W^\pm$ and Z bosons, but experiments have shown that their masses are non-zero [14, 15]. The mass of the $W^\pm$ is 80.379 ± 0.012 GeV and the Z mass has a value of 91.1876 ± 0.0021 GeV [11]. To solve this apparent contradiction, a theory of spontaneous symmetry breaking has been developed independently by Higgs [8] and Englert and Brout [9]. The particles obtain their masses through this broken symmetry. It is introduced by a complex scalar field ϕ that can be written as a weak isospin doublet with two complex scalar fields. The potential

$$V(\phi) = \mu^2(\phi^*\phi) + \lambda(\phi^*\phi)^2, \tag{2.1}$$

[1] L stands for left handed chiral particles and Y for the weak hypercharge

also known as the Higgs potential, is added to the corresponding Lagrangian $\mathcal{L}$. The potential has a finite minimum if λ is greater than zero. If μ^2 is also positive, the potential $V(\phi)$ has a minimum at $\phi = 0$ and the vacuum state corresponds to the scalar field ϕ being zero. The shape of the potential $V(\phi)$ and its minimum is illustrated in Figure 2.1(a). But if $\mu^2 < 0$, the potential takes on a different shape where the minimum is at $v = \pm\sqrt{-\frac{\mu^2}{\lambda}} \neq 0$, the vacuum expectation value. This is shown in Figure 2.1(b). The parameter v has a value of 246 GeV, and the masses of the gauge bosons and the Higgs boson depend on v [11]:

$$m_H = \sqrt{2\lambda}v, \tag{2.2}$$

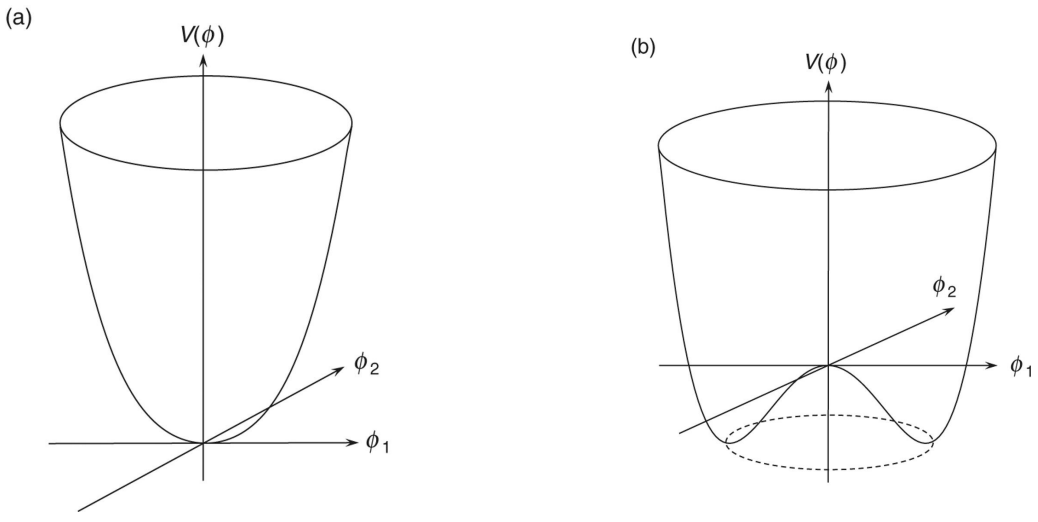
$$m_W = \frac{1}{2}g_W v, \tag{2.3}$$

$$m_Z = \frac{1}{2}\frac{g_W v}{\cos\theta_W}. \tag{2.4}$$

Here, g_W is the weak coupling constant. The mass of the Higgs particle has been measured by the Atlas and Cms experiments in 2012 [16, 17]. The latest measured value is 125.25 ± 0.17 GeV [11]. The value of parameter λ follows from the measurements of the Higgs mass (Equation 2.2) and has a value of $\lambda \simeq 0.13$ [11]. The resulting Lagrangian [18]

$$\mathcal{L} = (D_\mu\phi)^* (D^\mu\phi) - V(\phi) \tag{2.5}$$

with $D_\mu = \partial_\mu + igB_\mu$ where B_μ describes the new gauge field and the Higgs potential $V(\phi)$ is given in Equation 2.1.



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

Figure 2.1.: The Higgs potential $V(\phi)$ from Equation 2.1 for different values of the parameter μ^2 [18].

2.3. Higgs boson properties

2.3.1. Higgs boson decay

The Higgs boson can decay to many particles in the SM. Branching ratios (BR) state the relative frequency of a specific decay mode. The main reason why some BRs are smaller than others is that the mass of the particles in the final state are proportional to the coupling strength to the Higgs field [18]. That means the decay to light fermions is suppressed which makes it difficult to observe in experiments. The most sensitive Higgs decay modes are $H \to b\bar{b}$, $H \to WW^*$ and $H \to \tau^+\tau^-$. The decay into two on-shell W bosons cannot occur because the mass of two W bosons are larger than the mass of the Higgs boson. This means one of the W bosons must be off-shell[2]. Additionally, the Higgs boson can decay into two gluons and photons although gluons and photons are massless and do not couple directly to the Higgs field. Therefore, they decay through loops of virtual massive W bosons and top quarks, as shown in Figure 2.2. The values of the Higgs boson BRs are summarised in Table 2.2. BRs of direct decays into a fully leptonic state or in a semileptonic state with quarks with a $BR \leq 2.5$ % have not been listed in this table. The decay into two photons $\gamma\gamma$ has a very low branching ratio of 0.2 % but due to the sensitivity of the final state this decay mode is investigated in the detector [19].

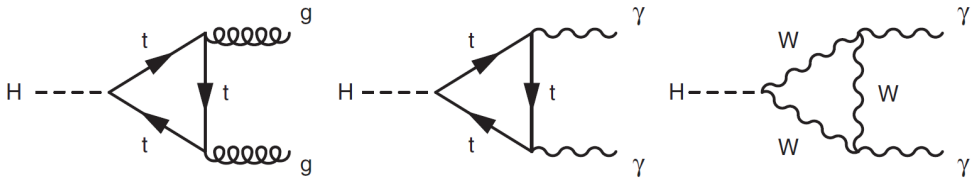


Figure 2.2.: The Feynman diagrams of the Higgs bosons decaying into two massless gluons and two massless photons [18].

[2]It is labelled with * and the same is true for the decay into two Z bosons.

Table 2.2.: The branching ratios for the different decay modes of the Higgs boson at a mass of $m_H = 125$ GeV [20].

Decay mode	Branching ratio
$H \to b\bar{b}$	58.2 %
$H \to WW^*$	21.4 %
$H \to gg$	8.9 %
$H \to \tau^+\tau^-$	6.3 %
$H \to c\bar{c}$	2.9 %
$H \to ZZ^*$	2.6 %
$H \to \gamma\gamma$	0.2 %

2.4. Higgs boson pair production

Higgs bosons can be produced in pairs at the LHC. Figure 2.3 shows the Feynman diagrams of the dominating Higgs pair production modes: the gluon-gluon fusion (ggF) mode and vector boson fusion (VBF) mode.

Figure 2.4 shows the cross-sections of the Higgs pair production at different centre of mass energies. The low values of the cross-section at centre of mass energies of $\sqrt{s} = 13$ TeV makes it difficult to observe the pair production in experiments. Measurements of the cross-section of Higgs boson pair production constrain the parameter λ which helps define the shape of the Higgs potential.

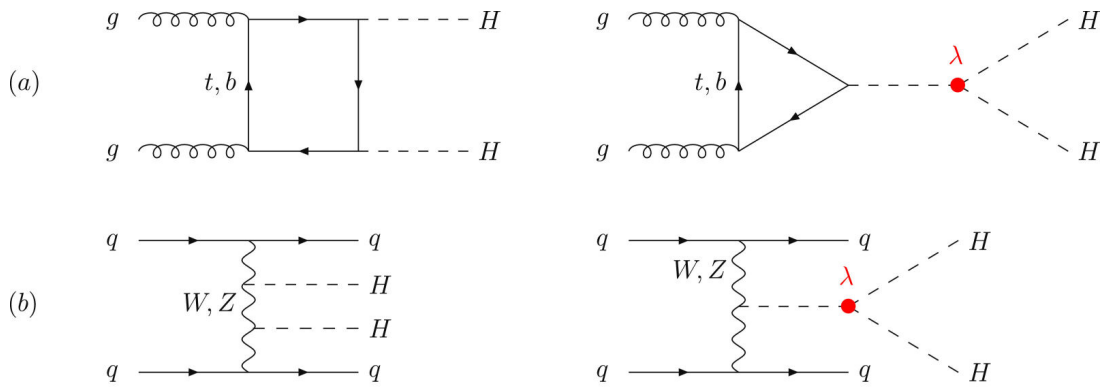


Figure 2.3.: The leading order (LO) Feynman diagrams for the Higgs pair production. The triple Higgs vertex is labelled with a red dot [21].

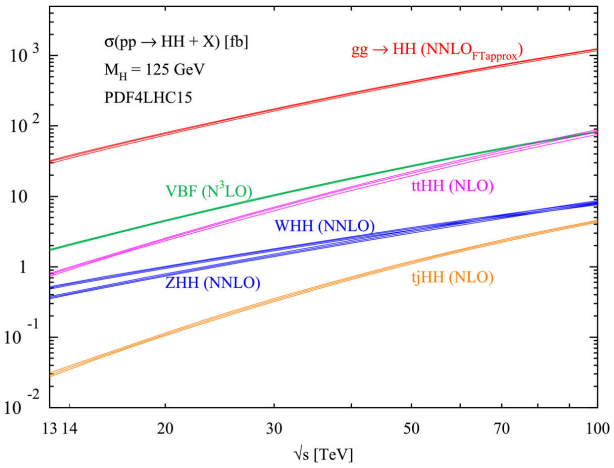


Figure 2.4.: The cross-sections of the Higgs pair production modes as a function of the centre of mass energy $\sqrt{s}$ in pp collisions [21].

2.4.1. Resonant Higgs Pair Production

Apart from the non-resonant Higgs pair production predicted by the SM, Higgs boson pairs can also be produced resonantly in beyond the SM theories which predict one or more additional heavy scalar particles. One example would be two-Higgs-doublet models [22] where h denotes a Higgs state with a different mass than H. The Feynman diagram in Figure 2.5 shows an example where h is produced via gluon fusion. Going one step further, models such as the two real singlet model (TRSM) [23] even predict $X \to SH$ where X and S are both heavy scalar bosons. None of these models predict a specific mass range for the new scalar bosons, and thus, searches must consider a wide mass range. Chapter 4 gives a specific insight to the TRSM.

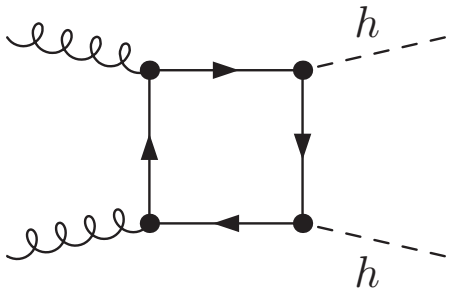


Figure 2.5.: The Feynman diagram for the Higgs production through gluon fusion at leading order in the two-Higgs-doublet model [24]. h represents Higgs boson pairs $h = h^0, A^0$ or H^0 in the two-Higgs-doublet model.

3. The Large Hadron Collider (LHC) and the ATLAS Detector

This chapter gives a overview of the ATLAS [6] detector at the Large Hadron Collider (LHC) [5] at the research centre CERN in Geneva.

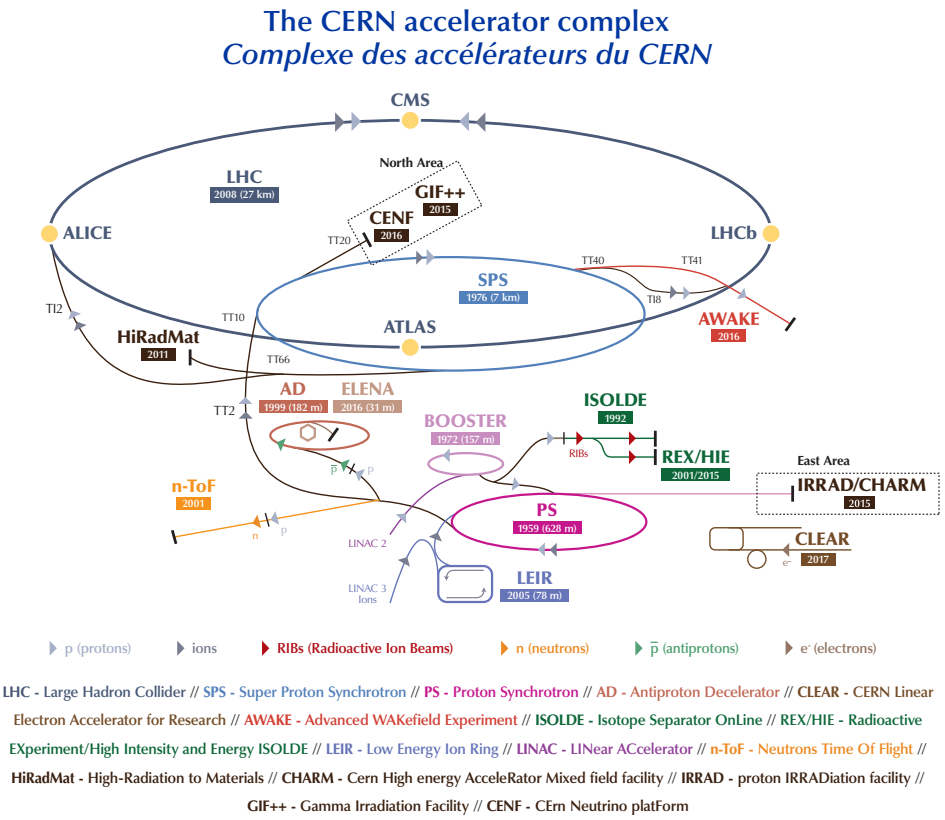


Figure 3.1.: An outline of CERN and its colliders and detectors [25].

3.1. The Large Hadron Collider (LHC)

In 2000, the Large Electron-Positron Collider (LEP) was closed to make place for a new hadron collider, the LHC. The LHC is a high-energy proton-proton (pp) collider with a circumference of 27 km and lies approximately 175 m under the border of France and Switzerland. An outline of the entire CERN complex with the LHC is shown in Figure 3.1.

With a centre of mass energy up to $\sqrt{s} = 14$ TeV, the LHC is the particle accelerator with the highest energy in the world. The LHC was build to test the SM and enable searches for theories beyond the SM, for example supersymmetry (SUSY).

Three runs, Run 1 ($\sqrt{s}$ = 7-8 TeV) from 2010 until 2013 [26], Run 2 ($\sqrt{s}$ = 13 TeV) between 2015 and 2018 [27] and the latest starting in 2022 ($\sqrt{s}$ = 13.6 TeV) have taken place. Between these runs, the shutdowns have been utilised to improve and upgrade the collider.

3.2. The ATLAS Detector

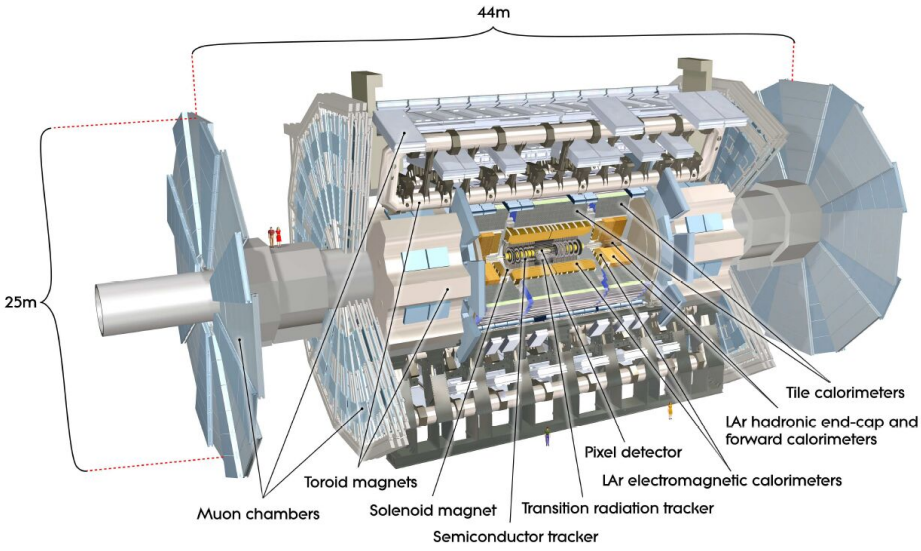


Figure 3.2.: The ATLAS detector and its buildup [6].

The ATLAS (**A** **T**oroidal LHC **A**pparatu**S**) [6] detector is a large general purpose detector of the LHC (along with the CMS detector). The setup of the detector is shown in Figure 3.2. The detector is 25 m in diameter and 44 m long and weighs about 7000 tons. pp collisions are examined in the detector. The centre of the detector is the collision point of the proton beams.

The origin of the right-handed coordinate system that is used at ATLAS corresponds to the interaction point of the collision. The beam direction defines the z-axis. The x-axis points to the centre of the LHC ring and the y-axis points upwards. Besides the cartesian coordinates, the spherical coordinate system is mainly used at ATLAS. The beam angle θ is the angle between the beam axis and the x-y-plane and the azimuthal angle ϕ is measured around the beam axis in the x-y-plane. A parameter that is also used, mainly for heavy particles, is the rapidity $y = \frac{1}{2} \ln [(E + p_z)/(E - p_z)]$ because of the invariance of rapidity differences Δy under a boost in beam direction. If the mass of the measured object is negligible compared to its energy, i.e. in the relativistic limit, y reduces to the pseudorapidity $\eta = -\ln \left[\tan \left(\frac{\theta}{2}\right)\right]$. The transverse momentum is defined by $p_T = \sqrt{p_x^2 + p_y^2}$. The transverse energy is defined as $E_T = \sqrt{(E_x^{\text{miss}})^2 + (E_y^{\text{miss}})^2} = \sqrt{m_x^2 + p_T^2}$. The missing transverse energy is defined as $E_T^{\text{miss}} = \sqrt{E_x^2 + E_y^2}$. It is used to calculate the energies of the invisible particles that are produced. ΔR is the distance in the pseudorapidity-azimuthal angle space and is defined as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

The ATLAS detector is composed of subsystems surrounding the interaction point of the pp collisions. The first detector that the collision products reach is the inner detector (ID), followed by the electromagnetic and hadron calorimeter. The muon spectrometer surrounds the entire detector. All detector parts are divided into a barrel and an end-cap.

3.2.1. The Inner Detector (ID)

The ID is the first detector system that particles emerging from LHC collisions reach. A solenoid magnetic field with a strength of 2 T immerses the inner detector which leads to curved trajectories of charged particles. The innermost layer consists of the pixel and strip (SCT) detectors which are semiconductor detectors. A transition radiation tracker (TRT) makes up the last part of the ID. These three components measure the transverse momentum p_T of charged particles and identify the vertex, the interaction point of two particles. The momentum resolution of the ID is $\sigma_{p_T}/p_T = 0.05\ \%\ p_T \oplus 1\ \%$ [GeV].

The pixel detector, with over 92 million silicon pixels, consists of four layers of modules. The first layer is the insertable b-layer that was installed in May 2014 and is the latest addition to the ID [28]. It was build to cope with high radiation. The pixels have an area of $50 \times 250\ \mu\text{m}^2$ in the innermost region and an area of $50 \times 400\ \mu\text{m}^2$ in the external layers.

The pixel detector is followed by the strip detector (SCT). The strips are placed onto modules and every strip is 80 μm wide and 250 μm long. The entire SCT has over 4000 modules with overall six million readout strips and every module consists of two layers of

strips. Eight strips are typically traversed by a particle. While the first layer is arranged parallel to the beam direction, the second layer is rotated by an angle of 40 mrad which is about 2.3° to measure two coordinates of a traversing particle. The pixel and strip detector cover a region of $|\eta| < 2.5$.

The TRT is the last component of the ID. It is mainly used to differentiate electrons from hadrons, mainly pions. The TRT consists of tubes (diameter of $d = 4$ mm) that act like drift chambers. The tubes are filled with $XeCO_2O_2$ gas where the molecules are ionised by the traversing particles. Additionally, photons are emitted and ionise the gas. Electrons that traverse the tubes liberate more photons than hadrons because the amount of the radiation depends on the mass of the particles. Heavy particles (hadrons for example) liberate more photons than light particles (electrons for example). The radiation also depends on the energy of the traversing particle. In general, electrons have lower energies than hadrons. Therefore, more photons are produced if a electron traverses the with gas filled area between the tubes. These are the reasons why traversing electrons have a higher transition radiation than hadrons and thus, the more photons are emitted, the more molecules in the tubes are ionised and therefore, more electrons reach the gold-plated wire inside each tube. The emitted electrons cause higher threshold hits than hadrons [29]. An electron or a hadron traverses about 36 tubes. The TRT covers a region of $|\eta| = 2.0$.

3.2.2. The Calorimeter System

The calorimeter of the ATLAS detector surrounds the ID and is used to measure the energy deposits that a particle leaves. Due to the energy loss, the particle loses momentum and stops in the calorimeter. During the interaction with the material, other particles are produced and create particle showers. The calorimeter system consists of two parts: the first one is the electromagnetic calorimeter for electrons, positrons and photons, and the second one is the hadron calorimeter which is designed to measure hadron energies.

The Electromagnetic Calorimeter

The electromagnetic calorimeter at ATLAS is a sampling calorimeter. Liquid argon (LAr) is the active material. To ensure that the argon remains in liquid form, the electromagnetic calorimeter has to be cooled down to a temperature of $-184°$ C. The absorber is made out of lead. The entire electromagnetic calorimeter has a thickness of over 22 radiation lengths in the barrel and over 24 radiation lengths in the end-cap region. The barrel region of the electromagnetic calorimeter covers the region with $|\eta| < 1.475$ and the end-cap

components a region of $1.375 < |\eta| < 3.2$. The energy resolution for the barrel and end-cap region of the electromagnetic calorimeter amounts to $\sigma_E/E = 10\ \%/\sqrt{E} \oplus 0.7\ \%$ [GeV] for electromagnetic showers.

The Hadronic Calorimeter

The hadronic calorimeter, which is also a sampling calorimeter, measures hadronic energies. Similar to the electromagnetic calorimeter, the hadronic calorimeter also has a barrel and end-cap region (interaction length of 9.7 in the barrel region and 10 interaction lengths in the end-cap region). The calorimeter for the hadrons is divided into three sections: a tile calorimeter (barrel and and two extended barrels), an end-cap calorimeter and a forward calorimeter. The energy resolution for hadron showers in the barrel and end-cap regions has a value of $\sigma_E/E = 50\ \%/\sqrt{E} \oplus 3\ \%$ [GeV].

For the tile calorimeter (3 layers for each barrel), scintillating tiles are used as the active material whereas steel is the material of the absorber. The barrel region covers $|\eta| < 1.0$ and the end caps range is $0.8 < |\eta| < 1.7$.

The second part of the hadron calorimeter is the LAr end-cap calorimeter on each side which is placed behind the barrel of the tile calorimeter. Tungsten, copper or lead is used as the passive material. The end-cap calorimeter covers a range of $1.5 < |\eta| < 3.2$ with four layers in total.

The third constituent of the hadronic calorimeter is the forward calorimeter ($3.1 < |\eta| < 4.9$, 3 layers). Similar to the end-cap calorimeter, LAr is the active material and tungsten, copper or lead is used as the passive material. Each end-cap consists of three modules. The energy resolution of the forward calorimeter differs with respect to the other constituents of the hadronic calorimeter and its value is $\sigma_E/E = 100\ \%/\sqrt{E} \oplus 10\ \%$ [GeV] for hadronic showers.

3.2.3. The Muon Spectrometer

The muon spectrometer is the outermost layer of the ATLAS detector and surrounds the ID and the calorimeters. The muon spectrometer measures the momentum of emitted muons that traverse through the entire detector by deflecting their trajectories. The muon spectrometer is surrounded by three toroid magnets (one in the barrel region and one for each end-cap region). The pseudorapidity region that the muon spectrometer covers is $|\eta| < 2.7$. The toroid magnet in the barrel region provides a bending power $\int \vec{B} \cdot d\vec{l}$ of 1.5 to 5.5 T·m and the magnet in the end-caps has a bending power of 1.5 to 7.5 T·m. At a transverse momentum of $p_T = 1$ TeV, the momentum resolution amounts

to 10 %.

The spectrometer is made out of gas filled chambers where the muons ionise the gas molecules. Signals can be then obtained and this gives an indication of the position of the muons. For different pseudorapidity ranges, different chambers with different granularities are utilised. The precise measurement of the muon coordinates is done with the Monitored Drift Tubes (MDT) for $|\eta| < 2.7$. The Cathode Strip Chambers (CSC) are used for tracking in a larger pseudorapidity region ($2.0 < |\eta| < 2.7$) due to their higher granularity.

Furthermore, a trigger system for $|\eta| < 2.4$ is used in addition to the MDT and CSC. The Resistive Plate Chambers (RPC) have been inserted for the barrel region and Thin Gap Chambers (TGC) are installed in the end-cap region.

3.2.4. The Trigger System

The trigger is used to filter the large amount of data produced by the LHC in real time [30]. The event rate at ATLAS is around 40 MHz at a luminosity of $\mathcal{L} = 10^{34}$ cm^{-2}s^{-1} and not every event is relevant. Thus, a two stage trigger is installed that decides if an event is kept or discarded.

The first component of the trigger is the Level-1 (L1) hardware-based trigger. It uses the information from the calorimeters and muon spectrometer to reduce the rate to approximately 100 kHz. It makes fast decisions but with low resolution and searches for particles with a large transverse momentum and energy. The High-Level-Trigger (HLT) is software-based and can make more complex decisions. It reduces the rate to around 1 kHz and uses information not only from the calorimeter and muon chambers, but also from the ID. Events passing the HLT are then saved for offline analysis.

4. The $HH/SH \to b\bar{b}VV^*$ decay channel with one lepton in the final state

A short introduction to the decay channel $X \to HH/SH \to b\bar{b}VV^* \to b\bar{b}q\bar{q}\ell\nu_\ell$ ($V = W^\pm,\ Z$) and the analysis will be given here. More aspects and further information about the analysis can be taken from Refs. [31, 32].

4.1. The $HH/SH \to b\bar{b}WW^*$ decay channel

Some theories beyond the Standard Model (BSM) predict a massive scalar boson X that can decay in either two Higgs bosons, or into one Higgs boson and an intermediate scalar boson S. These theories do not predict the masses of the X and S bosons. Therefore analyses must consider multiple mass hypotheses and mass ranges for the X and S bosons are scanned. The ranges for m_X and m_S that are considered in this analysis are:

$$750 \text{ GeV} \leq m_X \leq 5000 \text{ GeV}, \tag{4.1}$$

$$170 \text{ GeV} \leq m_S \leq 2500 \text{ GeV}. \tag{4.2}$$

Due to the criteria

$$m_X - m_S > m_H, \tag{4.3}$$

the Higgs and S bosons are produced on-shell. One Higgs boson decays into a $b\bar{b}$ pair and the decay products of the S (or the Higgs boson) are two vector bosons VV^*. This thesis will focus on the vector bosons being W bosons. Because only $m_S \geq 170$ GeV is considered, the two W bosons are on-shell. In the one lepton final state, one W boson decays into a $q\bar{q}$ pair and is named the W_{had} boson, and the other one decays into a lepton and the corresponding neutrino. This thesis concentrates on the lepton being a muon due to the better identification in a boosted topology (see next paragraph) compared to the

identification of an electron. The decay channel is sketched in Figure 4.1 for $X \to SH$ and $X \to HH$.

Due to the high momenta of the decay products, the topology of the final state is known as the boosted topology. This means that the $b\bar{b}$ and $q\bar{q}$ pairs cannot be resolved, and the direction of the lepton and the decay products of W_{had} do not span a large angle in the detector. The directions of the lepton and W_{had} are almost identical. Hence, in this analysis the boosted topology is defined by the angular separation $\Delta R(W_{\text{had}}, \ell)$ of the lepton and W_{had} boson being smaller than 1.

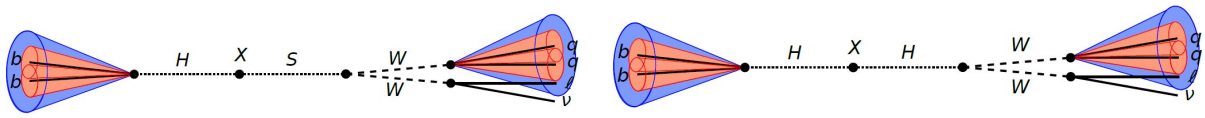


Figure 4.1.: The decay channel $X \to HH/SH \to b\bar{b}WW^* \to b\bar{b}qq\ell^-\overline{\nu}_\ell$ [32]. The single jets coming from the decay products are marked as red cones and the TAR jets are the blue cones.

4.2. Background Processes

Apart from the signal process displayed in Figure 4.1, various other processes can end up in the same final state by either having the same final state particles or misidentifying objects in the reconstruction. While the focus of this thesis is put on the W+jets background, other background processes must be taken into account as well:

- di-boson (two heavy gauge bosons are produced),
- non-prompt leptons (produced in later (mainly hadron) decays; dominated by QCD processes),
- single top,
- $t\bar{t}$,
- Z+jets.

Except the data-driven non-prompt background, all background processes are simulated by Monte Carlo (MC) methods [33, 34].

4.3. Object Reconstruction

All the decay products in the final state of the decay channel need to be reconstructed: the muon and the jets that come from the quark pairs in the final state. Due to the neutrino being invisible in the detector, it cannot be reconstructed. Therefore, it must be reconstructed indirectly.

4.3.1. Lepton Reconstruction

As mentioned before, this thesis will focus on the lepton being a muon μ. It is reconstructed by matching the track that it leaves in the ID with a track in the muon spectrometer [35]. The muon candidate must satisfy [31, 32]:

- a transverse momentum of $p_T > 10$ GeV,
- a pseudorapidity of $|\eta| < 2.5$ and
- the *loose* identification [36] criteria must be fulfilled. It ensures a high efficiency in the reconstruction of the muon and good reconstructed muon tracks.

Signal muons must also satisfy

- the *medium* [36] identification selection such that the systematic uncertainties that comes with the reconstruction of muon tracks is small and
- the *FixedCutTightTrackOnly* [37] isolation requirement. Muons are normally produced isolated from other particles and this isolation criteria requires that the scalar sum of all transverse momenta $p_T > 1$ GeV in a cone of $\min(10 \text{ GeV}/p_T^\mu, 0.3)$ around the muon with a transverse momentum of p_T^μ must be less than 6 % of p_T^μ [35].

4.3.2. Jet Reconstruction

In the final state of the decay channel, a $b\bar{b}$ and a $q\bar{q}$ pair is present. Due to colour confinement (described in Section 2.1), they do not propagate freely. The energy that is stored in the field between a quark and an antiquark increases as they separate in the detector. Therefore, a new quark-antiquark pair occurs. The quarks and antiquarks form colourless hadrons. This process continues and all hadrons produced are then reconstructed as a jet. For this, jet reconstruction algorithms have been derived taking into account the

distances [38]

$$d_{ij} = \min\left\{p_{T_i}^{2\rho}, p_{T_j}^{2\rho}\right\} \frac{\Delta R_{ij}^2}{R^2} \text{ and} \tag{4.4}$$

$$d_{iB} = p_{T_i}^{2\rho} \tag{4.5}$$

between two input objects i and j which can be calorimeter clusters, tracks or also other jets as well as the beam axis B. ΔR is the distance in the pseudorapidity-azimuthal angle space and R the fixed radius parameter of the cone as introduced in Section 3.2. If $d_{ij} < d_{iB}$, then the object i and j are merged together. But if $d_{ij} > d_{iB}$, object i will be identified as a jet and will no longer be considered in the jet reconstruction algorithm. These steps continue until all objects are assigned to jets. The exponent ρ adjusts the ordering of the energy scales compared to the geometrical scales.

This analysis makes use of the anti-k_T algorithm [38], an iterative cone algorithm with $\rho = -1$ and a radius parameter of $R = 0.4$, to cluster calorimeter deposits into jets. The anti-k_T algorithm focuses on grouping the highest energy clusters first and the less energetic objects afterwards.

TAR Jets

Track-Assisted Reclustered (TAR) [39] jets with a size-parameter of $R = 0.75$ are used (blue cones in Figure 4.1) in this analysis. Due to the large momenta of the Higgs/scalar bosons, their decay products are so close together that the particle flow jets (sketched as red cones; see next paragraph) overlap and the hadrons cannot be resolved individually. In the TAR jets, $R = 0.2$ jets are reclustered with the anti-k_T algorithm. All decay products are collected within a single jet. Additionally, they are required to have a transverse momentum of $p_T > 100$ GeV and a pseudorapidity $|\eta| < 2.0$.

The red cones in Figure 4.1 illustrate the $R = 0.4$ particle flow (PFlow) jets [40] originating from the $b\bar{b}$ pair and the two quarks $q\bar{q}$. Reconstructing the PFlow jets, information from the calorimeter and tracker are used to form the signal that should be identified with individual particles. Then, the jets are reconstructed with the anti-k_T algorithm. The red cones in Figure 4.1 encompasses the emitted hadrons which form a jet. These jets are required to have a $p_T > 20$ GeV and $|\eta| < 4.5$.

4.3.3. W_{had} and $H \to b\bar{b}$ Classification

In this analysis, two TAR jets are expected in this event, one originating from the Higgs boson that decays into $b\bar{b}$ ($H \to b\bar{b}$) and the other from the hadronically decaying W_{had}

boson (W_{had}). To distinguish the jets, a classifier is applied:

→ The TAR jet that has the smallest distance to the signal lepton is labelled W_{had} and the p_T leading jet that is not the W_{had} candidate is classified as $H \to b\bar{b}$.

4.3.4. b-Tagging

Due to the $b\bar{b}$ pair in this analysis, hadrons that contain b quarks appear in the jets. Hadrons consisting of b quarks have a relatively long lifetime compared to hadrons with different quarks as their constituents. As the hadrons are boosted, they travel a non-negligible distance in the detector before they decay into other particles. Hence, the primary vertex as the point of the pp collision can be distinguished from the secondary vertex which is the point of the hadron decay that contains at least one b quark. The secondary vertex gives information about the quark flavour (b quark) of the TAR jet [18].

At the ATLAS experiment, algorithms have been developed to tag jets containing b quarks [41, 42]. Low-level algorithms use two methods: the first method uses the reconstructed tracks of charged particles that the particles leave in the detector. The particle tracks then are combined with tracks coming from jets/hadrons. The second method utilises the long lifetime of hadrons that contain b quarks and with that the secondary vertex having a significant distance to the primary vertex. The secondary vertex gets reconstructed directly by combining the particle tracks. To maximise the efficiency of the low-level algorithm, high-level algorithms are used which, in this analysis, is the DL1r [43] algorithm, a deep neural network. It uses the information from the low-level algorithms to determine a score which corresponds to the probability of a certain jet type which means that the jet can either be b or c tagged or even a light jet can correspond to the output. The higher the output, the more likely it is to have a b jet.

4.4. Event Selection

4.4.1. Preselection

To ensure that the chosen events are placed in the boosted 1-lepton phase space, events are selected according to preselection criteria. For an event to pass the preselection, it must satisfy [31]:

- 1 signal muon must be present,
- at least two TAR jets (one from hadronically decaying W_{had} boson, the other one from $H \to b\bar{b}$) must be present in the event,

4. The $HH/SH \to b\bar{b}VV^*$ decay channel with one lepton in the final state

- $\Delta R(W_{\text{had}}, \ell) < 1$ with $\Delta R(W_{\text{had}}, \ell)$ being the angular separation between the hadronically decaying W_{had} boson and the lepton ℓ,
- $p_T > 500$ GeV for the $H \to b\bar{b}$ candidate and
- ≤ 2 b-tagged PFlow jets.

4.4.2. Region Definition

Signal regions (SR), a validation region (VR) and control regions (CR) are defined through cuts in order to distinguish between signal and background. Additionally, the regions are orthogonal to each other which means that no event can be found in more than one region. The cuts are chosen on variables that have different distributions for signal and each background process:

- number of b-tagged TAR jets in the event,
- the TAR jet mass $m_{\text{TAR}}^{H \to b\bar{b}}$ and $m_{\text{TAR}}^{W_{\text{had}}}$ coming from $H \to b\bar{b}$ and W_{had} candidates,
- the transverse mass $m_T^{W_{\text{lep}}}$ of the leptonically decaying W_{lep} boson (includes the muon and the missing transverse energy) and
- the substructure variable C_2 that checks if the $H \to b\bar{b}$ candidate is more consistent with the 2-prong signature rather a 1-prong one. C_2 is defined as [31, 32]

$$C_2 = \frac{ECF_3 \times ECF_2}{ECF_1^2}. \tag{4.6}$$

  with the Energy Correlator Functions (ECFs) that indicate the energy distributions inside a jet.

4.4.3. Control Regions

In order to control and correct the estimated background contributions, two CRs are defined for the two dominant background processes which are W+jets and $t\bar{t}$. In addition, a non-prompt CR is introduced in order to estimate the contributions of multijet events using data-driven methods. Different cuts define the three CRs. These are derived from the distributions of the variables shown in Figure 4.3. For events in the W+jets CR, no b-tagged TAR jet is required as shown in Figure 4.3(a). Furthermore, for an event candidate that is placed in the W+jets CR, the transverse mass $m_T^{W_{\text{lep}}}$ of the leptonically decaying W_{lep} boson must be between 60 GeV and 120 GeV due to the peak of $m_T^{W_{\text{lep}}}$ in

this interval for the W+jets background, plotted in Figure 4.3(b). $m_{\mathrm{TAR}}^{\mathrm{W_{had}}}$ is introduced to distinguish between the $t\bar{t}$ and W+jets, hence, it is only looked at in the CR of $t\bar{t}$ (Figure 4.3(c)).

Additionally, window cuts around the respective median of variables are introduced to distinguish between signal and background. For example, a m_H mass window around the Higgs mass $m_H = 125\,\mathrm{GeV}$ as the median is established because the $H \to b\bar{b}$ candidate peaks at 125 GeV for signal which can be seen in Figure 4.2. The distribution of the $H \to b\bar{b}$ mass for signal and the background is shown here. The event candidates only passed the preselection criteria. The m_H window has deviations depending on the boost of the Higgs boson and thus the collimation of the b quarks. An event saved in a specific region passes or fails the window. Events saved in one of the SRs pass it in order to fill the region with signal events, events saved in CRs fail it to enrich it with the selected background process. The smaller the m_H window, the more precise the selection of events for a region is if an event candidate stored passes the m_H window.

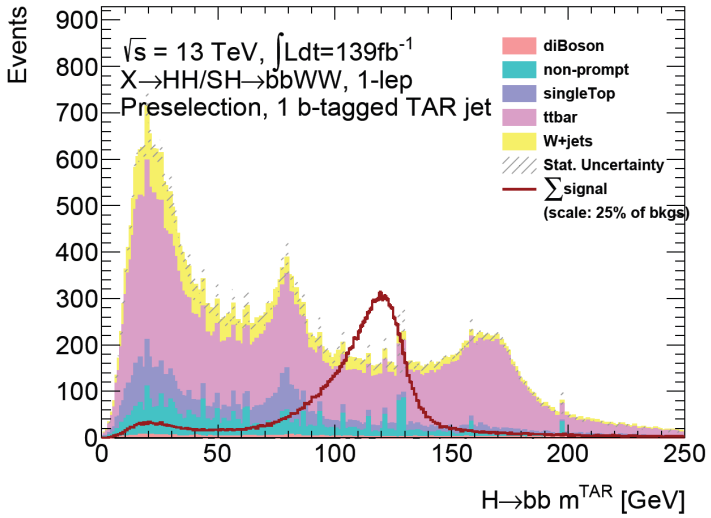
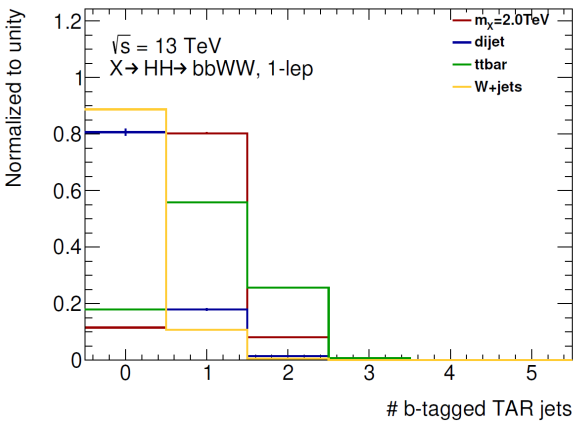


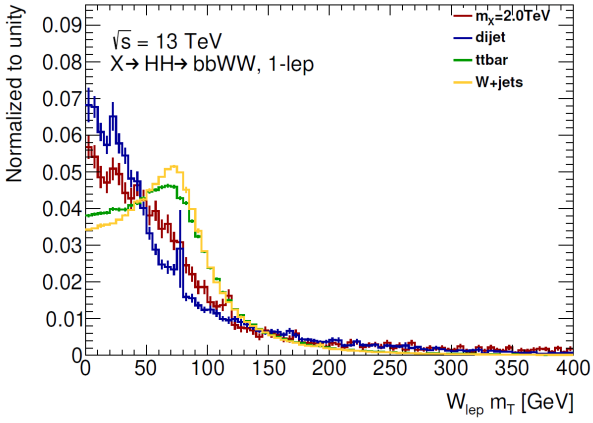
Figure 4.2.: The distribution of the $H \to b\bar{b}$ mass $m_{\mathrm{TAR}}^{H\to b\bar{b}}$ for signal and background [31]. The distribution of the signal takes on its maximum around the Higgs mass, the distribution of the background does not peak around the Higgs mass.

An event candidate that should be stored in the W+jets CR must not peak around the Higgs mass $m_H = 125\,\mathrm{GeV}$. This criteria follows from the distribution displayed in Figure 4.3(d). Here, the window cut is chosen at a 70 % signal efficiency which keeps a large amount of signal events but also excludes a decent amount of background events. For a $H \to b\bar{b}$ candidate to be stored in the W+jets CR, it must fail the 70 % m_H window.

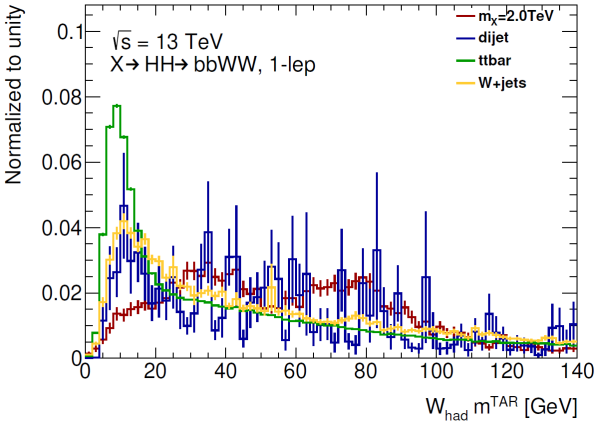
The consequential cuts for all three CRs are shown in Table 4.1.



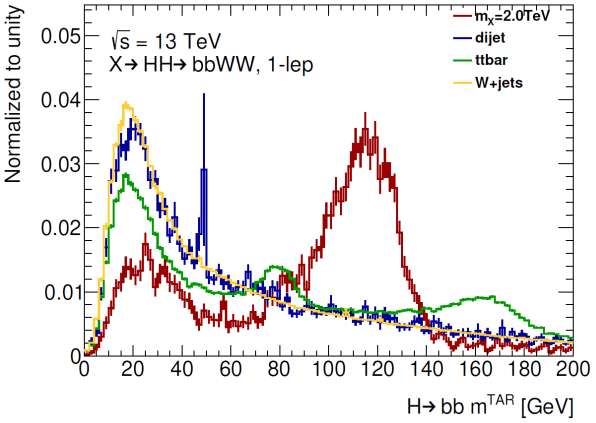
(a) The number of b-tagged TAR jets in the event for signal and background [31].



(b) The distribution of $m_T^{W_{\text{lep}}}$ for signal and background [31].



(c) The distribution of $m_{\text{TAR}}^{\text{W}_{\text{had}}}$ for signal and background [31].



(d) The distribution of $m_{\text{TAR}}^{H\to b\bar{b}}$ for signal and background [31].

Figure 4.3.: The distributions of the variables used to define the three CRs. Additionaly, the behaviour of the variables is shown in red for the signal ($m_X = 2$ TeV). The dijet MC approximates the non-prompt background [31].

Background Process	CR Cuts
W+jets	0 b-tagged TAR jets; $H \to b\bar{b}$ fails 70 % m_H window; 60 GeV $< m_T^{W_{\text{lep}}} <$ 120 GeV
QCD	0 b-tagged TAR jets; $H \to b\bar{b}$ fails 70 % m_H window; $m_T^{W_{\text{lep}}} <$ 60 GeV or $m_T^{W_{\text{lep}}} >$ 120 GeV
$t\bar{t}$	2 b-tagged TAR jets; $m_{\text{TAR}}^{W_{\text{had}}} <$ 20 GeV

Table 4.1.: The Control Region (CR) cuts for the three dominant background processes in this analysis.

In this analysis, two different W+jets CR will be examined: the first one is the CR that is described in Table 4.1 and is referred to as the W+jets CR1. For SHERPA 2.2.11, the m_H cut will be removed and this CR is denoted as the W+jets CR2.

4.4.4. Validation Region

For the VR, different variables define this region. The VR is established to cross-check the background estimation obtained from the CRs against data. The VR will be different for the SH and HH signal. The cuts for the VR are:

- 1 b-tagged TAR jet in the event,
- $H \to b\bar{b}$ candidate is single b-tagged and
- $H \to b\bar{b}$ candidate fails the 80 % m_H mass window (Figure 4.2).

The additional cuts for the SH and HH selection are:

- *SH selection:* $H \to b\bar{b}$ candidate passes 80 % C_2 window. The distribution is exhibited in Figure 4.4. The signal peaks around $H \to b\bar{b}$ $C_2 \approx 0.05$ and the distribution for the background processes have their maximum in the range $H \to b\bar{b}$ $C_2 \in [0.05, 0.25]$. Due to the small overlap of these peaks, the signal efficiency for the $H \to b\bar{b}$ C_2 window is therefore set at 80 %.
- *HH selection:* event candidate passes the SH selection criteria and passes 80 % $\Delta R(W_{\text{had}}, \ell)$ window. The distributions for signal and background are plotted in Figure 4.5. The widely ranged distribution for the background processes only has low values for small $\Delta R(W_{\text{had}}, \ell)$'s, where the distribution for the signal has its maximum. Thus, the signal efficiency for the $\Delta R(W_{\text{had}}, \ell)$ window is set at 80 %.

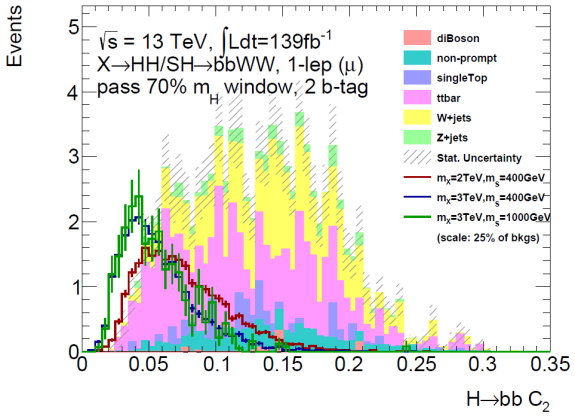


Figure 4.4.: The distributions of $H \to b\bar{b}\ C_2$ the SH selection [32].

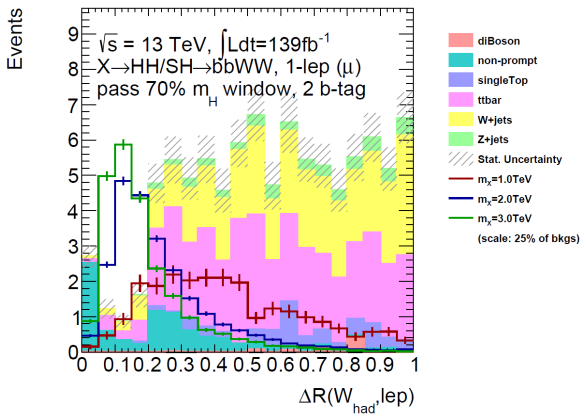


Figure 4.5.: The distributions of $\Delta R(W_{\text{had}}, \ell)$ for the HH selection [32].

For studying the visible+met mass $m_{\text{vis+met}}^{SH/HH}$ distribution (introduced in Section 5.1) as the most important variable in this thesis, the cuts on $H \to b\bar{b}\ C_2$ and $\Delta R(W_{\text{had}}, \ell)$ windows will be removed to increase the statistics. This specific VR will be referred to as VR_{spec}.

The criteria for the SRs can be taken from Refs. [31, 32].

4.4.5. Normalisation Factors

Additionally, normalisation factors are applied in the CRs and VR since a significant mismatch between the overall number of data and background events have been observed for W+jets, $t\bar{t}$ and non-prompt lepton estimate. During the course of this thesis, the W+jets MC generator has been updated to a new version (Sherpa 2.2.1 to Sherpa 2.2.11 [44]), such that two sets of normalisation factors have been derived and are listed in Table 4.2.

	W+jets	non-prompt	$t\bar{t}$
Sherpa 2.2.1	0.53	1.18	0.71
Sherpa 2.2.11	1.07	0.86	0.78

Table 4.2.: The normalisation factors for the three main background processes [31, 32].

5. Modelling in the W+jets Control Regions

In this chapter, the W+jets background modelling in the W+jets CR1 and W+jets CR2 (both CRs introduced in Section 4.4.3) will be analysed. For the background modelling, the MC simulations from SHERPA version 2.2.1 is used. The expected W+jets background will be compared to the observed data from which the other background processes (Section 4.2) are subtracted. Given the disagreement between data and background that is observed, the unweighted W+jets distributions are weighted based on different variables to then have a better agreement with data in control regions and validation regions (see Chapter 6). Additionally, the effect of the new SHERPA version 2.2.11 on the modelling is investigated.

5.1. Distributions of Variables in the W+jets CRs

Here, different variables are investigated for their modelling:

- $\Delta R(W_{\text{had}}, \ell)$,
- the missing transverse energy E_T^{miss},
- the TAR jet mass $m_{\text{TAR}}^{H\to b\bar{b}}$ of the $H \to b\bar{b}$ candidate,
- lepton p_T and
- the visible+met mass $m_{\text{vis+met}}^{SH/HH} = \sqrt{\left(p^{H\to b\bar{b}} + p^{\text{W}_{\text{had}}} + p^{\ell} + p^{\text{met}}\right)^2}$ with $p^{\text{met}} = (E_T^{\text{miss}}, p_x^{\text{miss}}, p_y^{\text{miss}}, 0)$. This variable is the most important one as it is the final discriminant used in the analysis, in the SR [32].

For the visible+met mass $m_{\text{vis+met}}^{SH/HH}$, the momentum four-vectors for the three final state objects (the TAR jet of $H \to b\bar{b}$, the hadronically decaying W_{had} and the lepton ℓ) are used. The missing momentum p^{met} is used to represent the neutrino kinematics, but with

p_z^ν set to zero. Other mass definitions (transverse mass and visible mass) lead to broader peaks in the signal distributions [32].

The distributions for the five variables in the W+jets CR1 and W+jets CR2 are shown in the Figures 5.1 and 5.2. As expected, the distribution in the W+jets CR1 and CR2 is dominated by the W+jets background process because the W+jets CR is enriched with the W+jets background process.

Generally, it can be said that the background modelling in the W+jets CR2 using Sherpa 2.2.11 improves drastically for the variables that do not exhibit a good agreement between data and background in the W+jets CR2 using Sherpa 2.2.1 - especially for the visible+met mass $m_{\text{vis+met}}^{SH/HH}$.

Only the distributions of $\Delta R(W_{\text{had}}, \ell)$ (Figure 5.1(a) and 5.1(b)) and lepton p_T (Figure 5.2(c) and 5.2(d)) exhibit a good agreement between data and MC simulation in the W+jets CR1 and W+jets CR2 and the statistical uncertainty of each data point covers the difference background and data. For the three other variables it can be seen that the observed data and background do not agree well for both Sherpa versions.

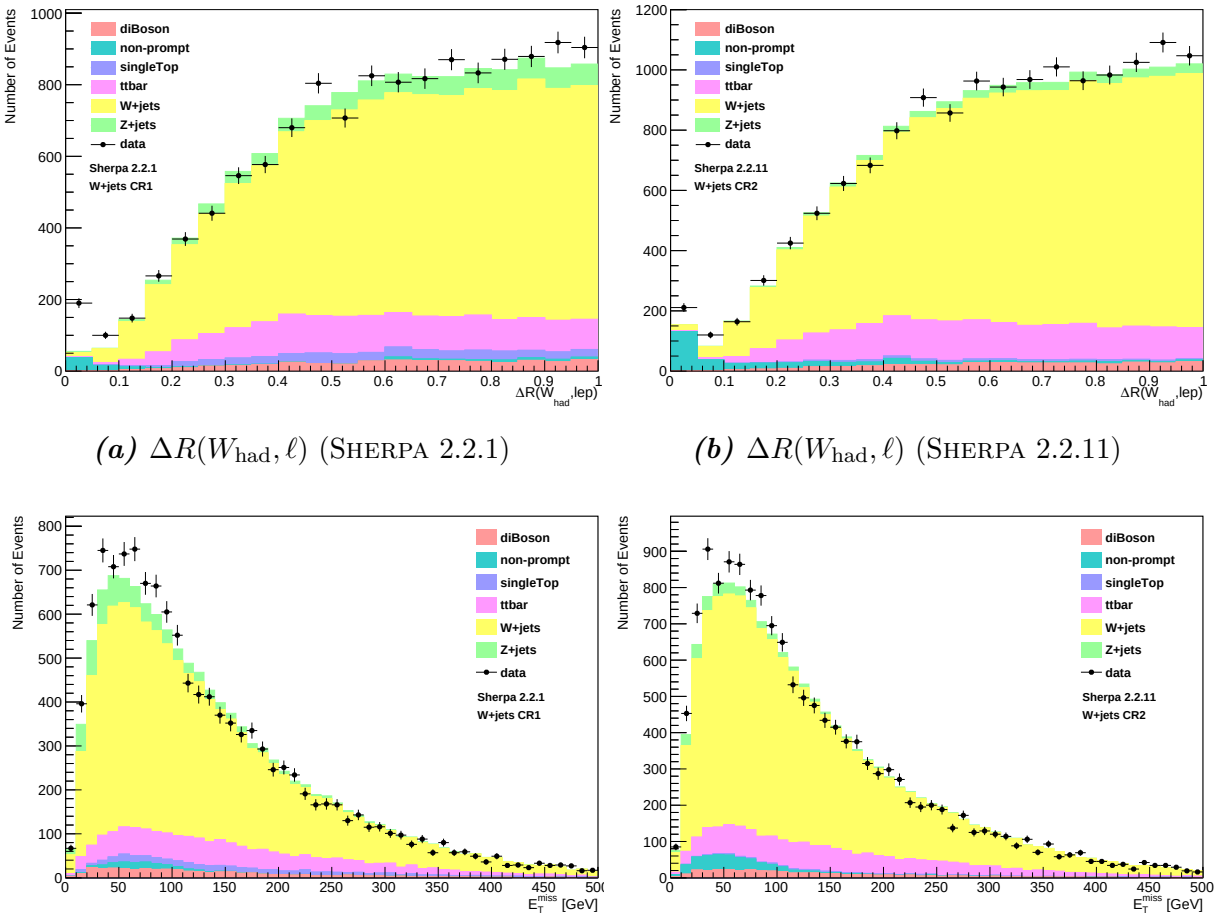
Around the peak at 50 GeV of the missing transverse energy E_T^{miss} distribution, an underestimation of background can be seen in the W+jets CR1 (Figure 5.1(c)) and W+jets CR2 (Figure 5.1(d)). In the tail of the E_T^{miss} distribution, the simulated background mostly overestimates the data in both W+jets CRs.

Figure 5.2(a) shows the background modelling (Sherpa 2.2.1) for $m_{\text{TAR}}^{H\to b\bar{b}}$. The cut in the range 100 GeV to 130 GeV comes from the definition of the W+jets CR (Table 4.1) and vanishes by removing the cut which can be seen in Figure 5.2(b) (Sherpa 2.2.11). For the background modelling in the W+jets CR1 of $m_{\text{TAR}}^{H\to b\bar{b}}$ in Figure 5.2(a), an underestimation of the background is observable for 135 GeV $< m_{\text{TAR}}^{H\to b\bar{b}} <$ 180 GeV. For the modelling of Sherpa 2.2.11 in the W+jets CR2 in Figure 5.2(b), the background underestimation does not occur and the background agrees with data. For masses lower than 100 GeV, the background modelling overestimates data in the W+jets CR1 but is covered by the systematic uncertainties. The modelling with Sherpa 2.2.11 in the W+jets CR2 exhibits the same features. For the mass range 45 GeV $\leq m_{\text{TAR}}^{H\to b\bar{b}} \leq$ 95 GeV, the new Sherpa sample underestimates the data more than before. But for the masses from 100 GeV onwards, the background modelling exhibits almost a total agreement with the data points.

For the visible+met mass $m_{\text{vis+met}}^{SH/HH}$, the mismodelling in the W+jets CR1 and W+jets CR2 in the Figures 5.2(e) and 5.2(f) is clearly notable. The background around the peak at 1500 GeV underestimates the data in almost every bin. For higher mass values (from 1900 GeV onwards), the background processes tend to overestimate data. However, the

background modelling using SHERPA 2.2.11 W+jets samples in the W+jets CR2 improves compared to the modelling using SHERPA 2.2.1 W+jets samples in the W+jets CR1.

The modelling of the W+jets background can be improved. By reweighting the W+jets background, such that an agreement between the background and data distribution in a specific variable is achieved.



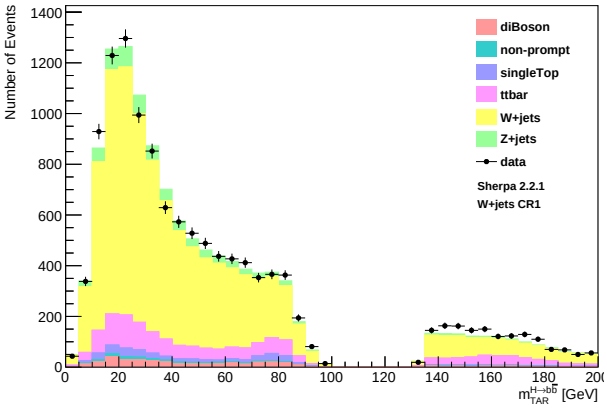
(a) $\Delta R(W_{\text{had}}, \ell)$ (SHERPA 2.2.1)

(b) $\Delta R(W_{\text{had}}, \ell)$ (SHERPA 2.2.11)

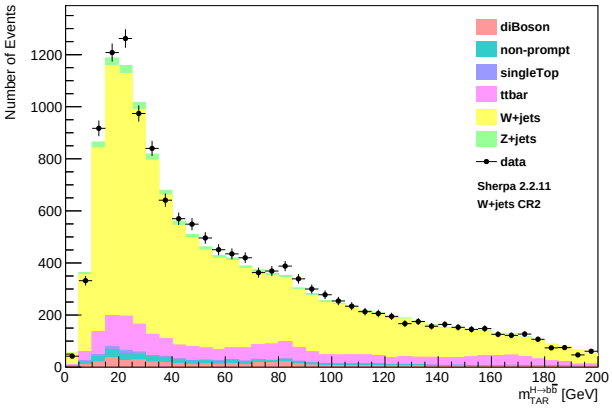
(c) E_T^{miss} (SHERPA 2.2.1)

(d) E_T^{miss} (SHERPA 2.2.11)

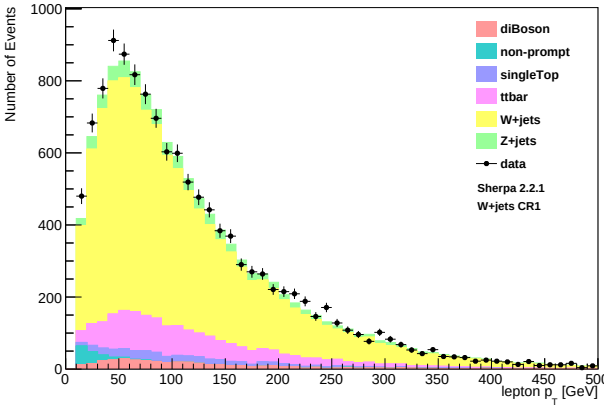
Figure 5.1.: Stacked histograms of $\Delta R(W_{\text{had}}, \ell)$ and E_T^{miss}. The first histogram of each variable exhibits the SHERPA 2.2.1 background modelling in the W+jets CR1, the second histogram of each variable the SHERPA 2.2.11 background modelling in the W+jets CR2.



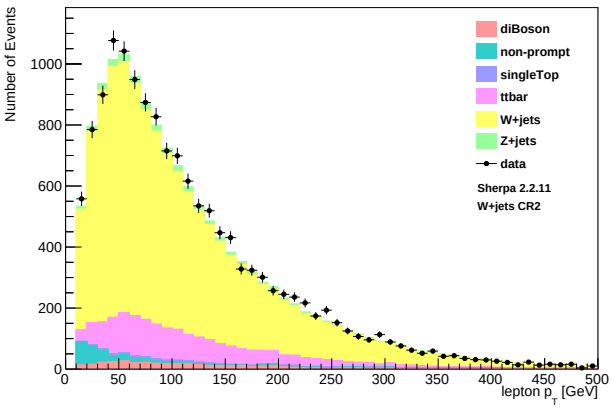
(a) $m_{\mathrm{TAR}}^{H\to b\bar{b}}$ (SHERPA 2.2.1)



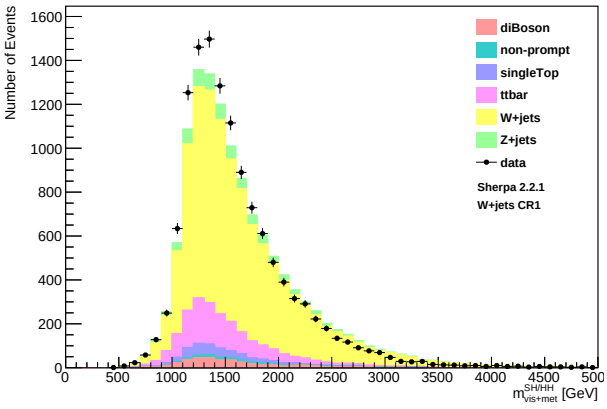
(b) $m_{\mathrm{TAR}}^{H\to b\bar{b}}$ (SHERPA 2.2.11)



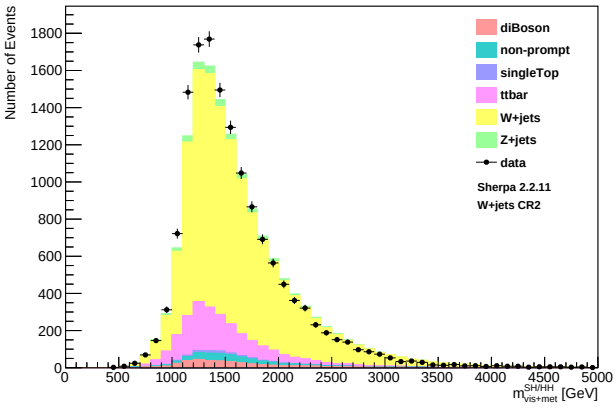
(c) lepton p_T (SHERPA 2.2.1)



(d) lepton p_T (SHERPA 2.2.11)



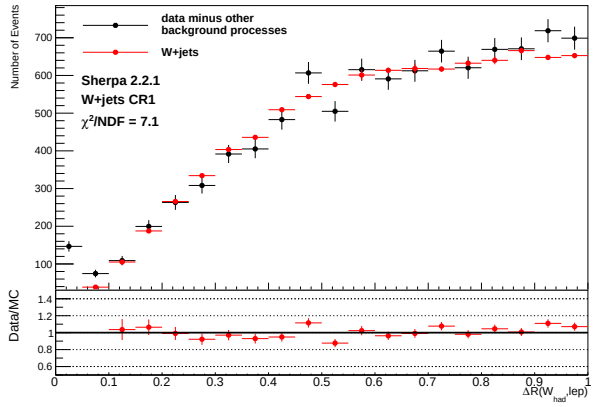
(e) $m_{\mathrm{vis+met}}^{SH/HH}$ (SHERPA 2.2.1)



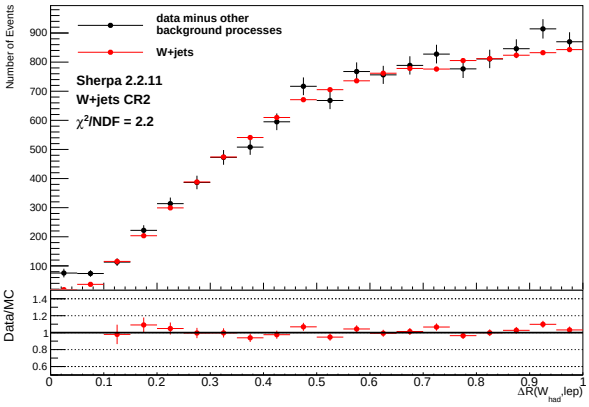
(f) $m_{\mathrm{vis+met}}^{SH/HH}$ (SHERPA 2.2.11)

Figure 5.2.: Stacked histograms of $m_{\mathrm{TAR}}^{H\to b\bar{b}}$, lepton p_T and $m_{\mathrm{vis+met}}^{SH/HH}$. The first histogram of each variable exhibits the SHERPA 2.2.1 background modelling in the W+jets CR1, the second histogram of each variable the SHERPA 2.2.11 background modelling in the W+jets CR2.

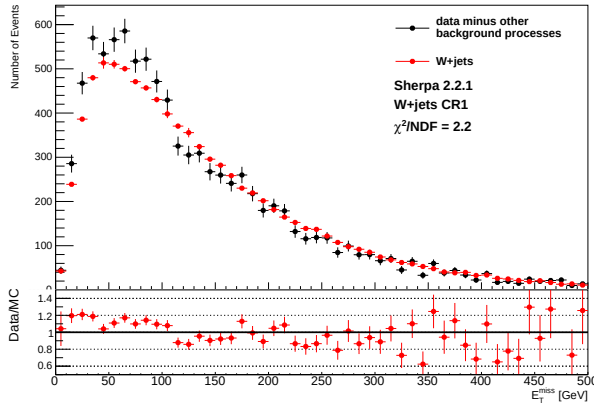
In the Figures 5.3 and 5.4, the number of events for the background prediction and data in the W+jets CR1 and W+jets CR2 is shown for the five different variables. To focus on the W+jets background ($\mathrm{MC}_{W+\mathrm{jets}}$), all other backgrounds are subtracted from data ($\mathrm{Data}-\mathrm{MC}_{\mathrm{else}}$). The mismodelling in the W+jets CR1 and W+jets CR2 for the above mentioned variables can be seen as well. In addition, the ratio $\frac{\mathrm{Data}-\mathrm{MC}_{\mathrm{else}}}{\mathrm{MC}_{W+\mathrm{jets}}}$ is plotted for each variable under their respective plot and denoted as "Data/MC" in the plots. If all ratio points are near one, then a good agreement between W+jets background and data is present.



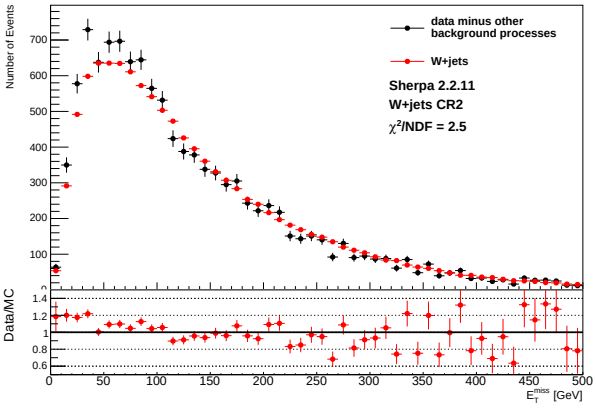
(a) $\Delta R(W_{\mathrm{had}}, \ell)$ (Sherpa 2.2.1)



(b) $\Delta R(W_{\mathrm{had}}, \ell)$ (Sherpa 2.2.11)

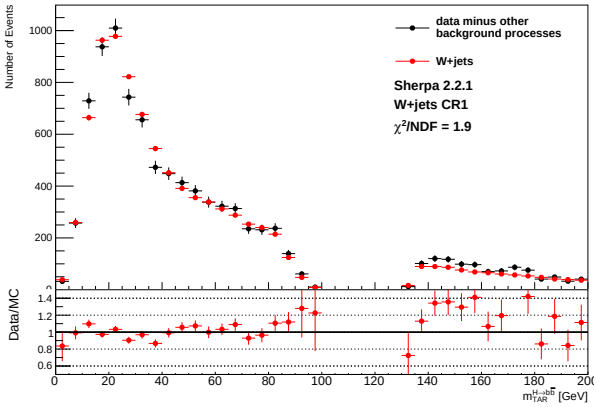


(c) E_T^{miss} (Sherpa 2.2.1)

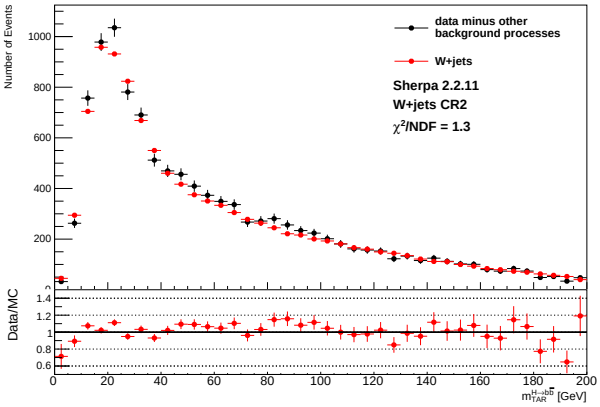


(d) E_T^{miss} (Sherpa 2.2.11)

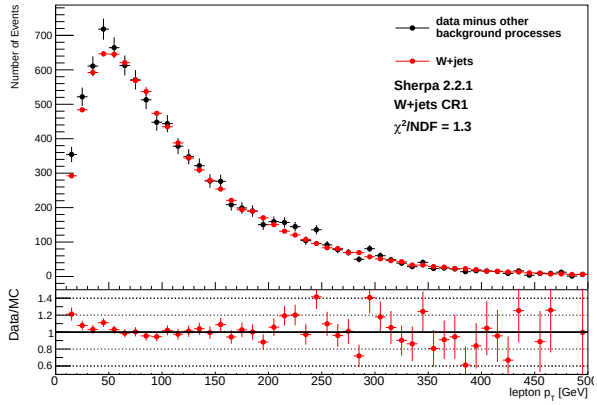
Figure 5.3.: The distributions of $\Delta R(W_{\mathrm{had}}, \ell)$ and E_T^{miss} for Sherpa 2.2.1 (left side) and Sherpa 2.2.11 (right side) W+jets and data, respectively of which other background processes have been subtracted.



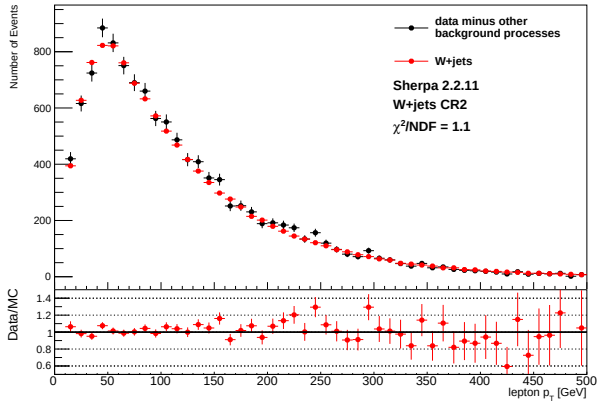
(a) $m_{\mathrm{TAR}}^{H\to b\bar{b}}$ (Sherpa 2.2.1)



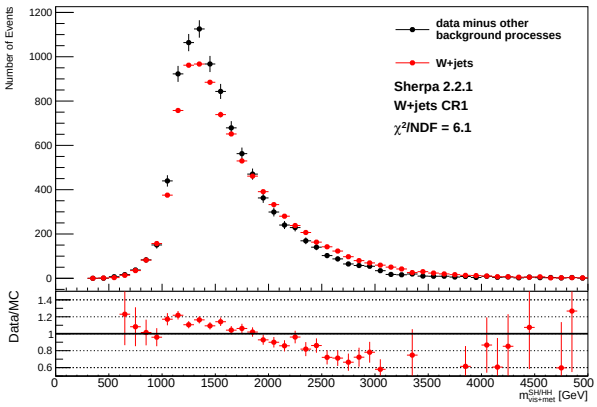
(b) $m_{\mathrm{TAR}}^{H\to b\bar{b}}$ (Sherpa 2.2.11)



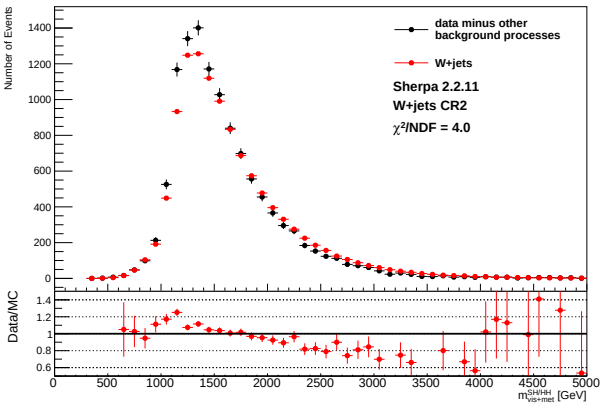
(c) lepton p_T (Sherpa 2.2.1)



(d) lepton p_T (Sherpa 2.2.11)



(e) $m_{\mathrm{vis+met}}^{SH/HH}$ (Sherpa 2.2.1)



(f) $m_{\mathrm{vis+met}}^{SH/HH}$ (Sherpa 2.2.11)

Figure 5.4.: The distributions of $m_{\mathrm{TAR}}^{H\to b\bar{b}}$, lepton p_T and $m_{\mathrm{vis+met}}^{SH/HH}$ for Sherpa 2.2.1 (left side) and Sherpa 2.2.11 (right side) W+jets and data, respectively of which other background processes have been subtracted.

5.1.1. Comparison of W+jets Background Modelling and Data via the χ^2/NDF Method

In order to compare data and the W+jets background and to see the possible improvement after reweighting the W+jets events, the χ^2 method is used. The χ^2 is defined as

$$\chi^2 = \sum_{i=1}^{\mathrm{NDF}} \frac{(x_i - \mu_i)^2}{\sigma_i^2}, \tag{5.1}$$

where x_i is the number of W+jets events in each bin i, μ_i is the number of data events in bin i, σ_i^2 is the variance of x_i and the sum goes from 1 to the number of degrees of freedom NDF $= N - 1$ which is defined as the number of bins N minus one. The larger χ^2/NDF is, the worse is the modelling of W+jets in the considered W+jets CR for the selected variable.

The values of χ^2/NDF for the nominal W+jets distribution are listed in Table 5.1 for Sherpa 2.2.1 and Sherpa 2.2.11 W+jets samples. The five values of χ^2/NDF underline the poor modelling of Sherpa 2.2.1 W+jets in the W+jets CR1, especially for $\Delta R(W_{\mathrm{had}}, \ell)$ and the visible+met mass $m_{\mathrm{vis+met}}^{SH/HH}$ with χ^2/NDF values of 7.1 and 6.1, respectively. Although, the W+jets distribution of $\Delta R(W_{\mathrm{had}}, \ell)$ in Figure 5.3(a) has a good agreement with data, the χ^2/NDF value is large. This is due to the small amount of W+jets events in the first bins.

For the χ^2/NDF values in the W+jets CR2, the values improve by a significant amount especially for $\Delta R(W_{\mathrm{had}}, \ell)$ and $m_{\mathrm{vis+met}}^{SH/HH}$. The entire background contribution in the first bin of the distribution using Sherpa 2.2.11 of $\Delta R(W_{\mathrm{had}}, \ell)$ is much larger in the W+jets CR2 than in the distribution using Sherpa 2.2.1, which explains the significant improvement of the χ^2/NDF value due to the better agreement with data in the first bin. The χ^2/NDF improves for $m_{\mathrm{TAR}}^{H\to b\bar{b}}$ and the lepton p_T whereas the χ^2/NDF value increases slightly for E_T^{miss}.

5.2. Improvement of the Background Modelling in the W+jets CR

5.2.1. Correlations between the Variables in the W+jets CRs

In order to see which distribution can be used to obtain a weight to apply to other distributions to improve the W+jets modelling in the respective W+jets CR, the linear correlations between the variables are determined. This is done by evaluating the dimen-

Variable	χ^2/NDF	
	SHERPA 2.2.1 (W+jets CR1)	SHERPA 2.2.11 (W+jets CR2)
$\Delta R(W_{\text{had}}, \ell)$	7.1	2.2
E_T^{miss}	2.2	2.5
$m_{\text{TAR}}^{H\to b\bar{b}}$	1.9	1.3
lepton p_T	1.3	1.1
$m_{\text{vis+met}}^{SH/HH}$	6.1	4.0

Table 5.1.: The values for χ^2/NDF for the nominal SHERPA 2.2.1 and SHERPA 2.2.11 W+jets background in the respective W+jets CR. The values have been obtained from the plots in Figure 5.3 and 5.4.

sionless correlation factor *corr* between two variables X and Y and their mean $\overline{X}$ and $\overline{Y}$ [45]:

$$corr = \frac{\sum_i^N \left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sqrt{\sum_i^N \left(X_i - \overline{X}\right)^2 \sum_i^N \left(Y_i - \overline{Y}\right)^2}} \tag{5.2}$$

A value of ± 1 implies that the two variables X and Y are fully correlated while a value of zero (or near zero) would indicate that these two variables are independent of each other. The value of the correlation factor predicts which ratio $\frac{\text{Data}-\text{MC}_{\text{else}}}{\text{MC}_{W+\text{jets}}}$ from a variables' distribution can be used to reweight the W+jets events in the distribution of another variable, so that data and background have a better agreement.

The correlation factors *corr* between the previously introduced variables using SHERPA 2.2.1 W+jets background in the W+jets CR1 are listed in Table 5.2 and for the SHERPA 2.2.11 W+jets background in the W+jets CR2 in Table 5.3.

Sherpa 2.2.1	$\Delta R(W_{\text{had}}, \ell)$	E_T^{miss}	$m_{\text{TAR}}^{H\to b\bar{b}}$	lepton p_T	$m_{\text{vis+met}}^{SH/HH}$
$\Delta R(W_{\text{had}}, \ell)$	**1**	−0.062	−0.013	−0.272	−0.058
E_T^{miss}	−0.062	**1**	0.029	−0.083	0.150
$m_{\text{TAR}}^{H\to b\bar{b}}$	−0.013	0.029	**1**	0.028	0.062
lepton p_T	−0.272	−0.083	0.028	**1**	0.130
$m_{\text{vis+met}}^{SH/HH}$	−0.058	0.150	0.062	0.130	**1**
$H \to b\bar{b}$ C_2	—	—	—	—	−0.043
$p_T^{H\to b\bar{b}}$	—	—	—	—	0.629
$m_{\text{TAR}}^{\text{W}_{\text{had}}}$	—	—	—	—	0.155
$p_T^{W_{\text{had}}}$	—	—	—	—	0.448
$m_T^{W_{\text{lep}}}$	—	—	—	—	0.009

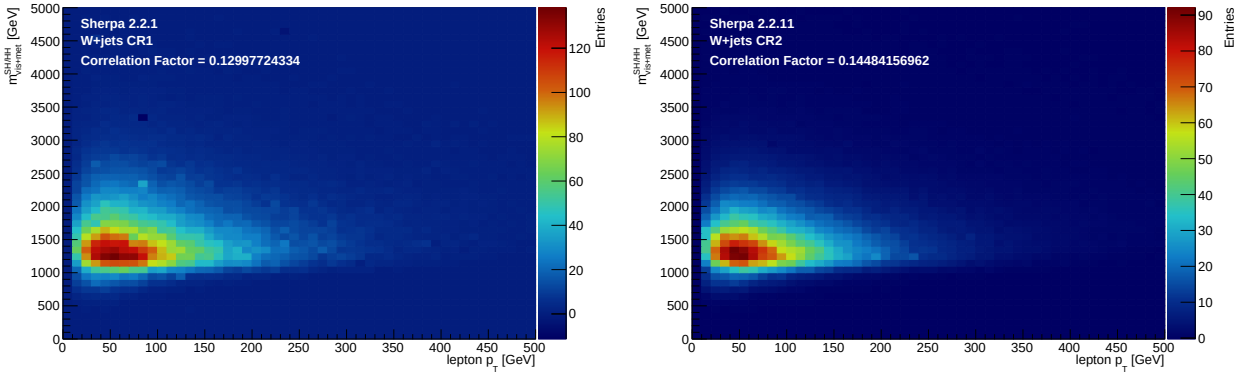
Table 5.2.: The correlation factors between the variables in the W+jets CR1. The correlation factors between the five additional variables (second part of the table) and the first four original variables (second to fifth column) are not computed because the five additional variables are only used to reweight the distribution of $m_{\text{vis+met}}^{SH/HH}$. The values are rounded to the third decimal place. The correlation plots can be found in Section A.1.

Sherpa 2.2.11	$\Delta R(W_{\text{had}}, \ell)$	E_T^{miss}	$m_{\text{TAR}}^{H\to b\bar{b}}$	lepton p_T	$m_{\text{vis+met}}^{SH/HH}$
$\Delta R(W_{\text{had}}, \ell)$	**1**	−0.071	−0.015	−0.257	−0.064
E_T^{miss}	−0.071	**1**	0.023	−0.075	0.152
$m_{\text{TAR}}^{H\to b\bar{b}}$	−0.015	0.023	**1**	0.026	0.046
lepton p_T	−0.257	−0.075	0.026	**1**	0.145
$m_{\text{vis+met}}^{SH/HH}$	−0.064	0.152	0.046	0.145	**1**
$H \to b\bar{b}$ C_2	—	—	—	—	−0.046
$p_T^{H\to b\bar{b}}$	—	—	—	—	0.620
$m_{\text{TAR}}^{\text{W}_{\text{had}}}$	—	—	—	—	0.116
$p_T^{W_{\text{had}}}$	—	—	—	—	0.429
$m_T^{W_{\text{lep}}}$	—	—	—	—	0.013

Table 5.3.: The correlation factors between the variables in the W+jets CR2. The correlation factors between the five additional variables (second part of the table) and the first four original variables (second to fifth column) are not computed because the five additional variables are only used to reweight the distribution of $m_{\text{vis+met}}^{SH/HH}$. The values are rounded to the third decimal place. The correlation plots can be found in Section A.1.

Overall, the correlation factors between the variables are very small. The only variables that have a non-vanishing ($|corr| > 0.1$) correlation with another variable are $\Delta R(W_{\text{had}}, \ell)$ with the lepton p_T, and E_T^{miss} and lepton p_T with the visible+met mass $m_{\text{vis+met}}^{SH/HH}$ in both W+jets CRs.

Besides these results, the correlation plots between these variables are produced and can be seen in Section A.1. The correlation plots between lepton p_T and $m_{\text{vis+met}}^{SH/HH}$ using Sherpa 2.2.1 and Sherpa 2.2.11 are shown in Figure 5.5 as an example.



(a) Sherpa 2.2.1 *(b)* Sherpa 2.2.11

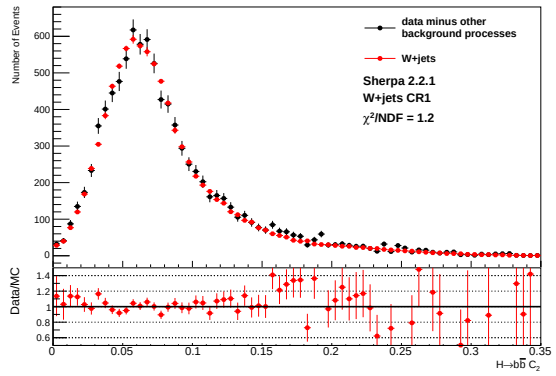
Figure 5.5.: The correlation plots between the lepton p_T and $m_{\text{vis+met}}^{SH/HH}$ for (a) Sherpa 2.2.1 and (b) Sherpa 2.2.11 W+jets background in their respective W+jets CR.

Due to the importance of the visible+met mass $m_{\text{vis+met}}^{SH/HH}$ as a discriminant in this analysis, five more variables are investigated with regards to use the distributions to obtain a weight for the $m_{\text{vis+met}}^{SH/HH}$ distribution:

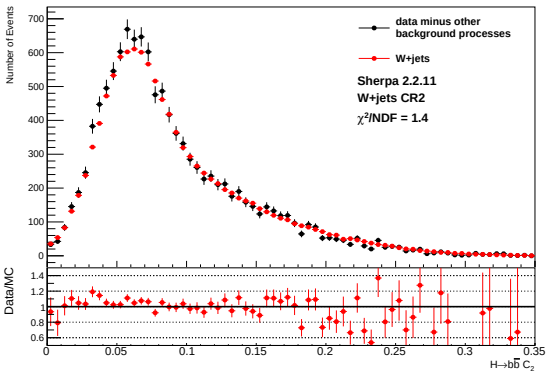
- $H \to b\bar{b}$ C_2, the substructure variable introduced in Section 4.4.2,
- $p_T^{H \to b\bar{b}}$, the transverse momentum of the $H \to b\bar{b}$ candidate,
- $m_{\text{TAR}}^{\text{W}_{\text{had}}}$, the mass of the decay products of the hadronically decaying W_{had} boson,
- $p_T^{W_{\text{had}}}$, the transverse momentum of the hadronically decaying W_{had} candidate and
- $m_T^{W_{\text{lep}}}$, the mass of the decay products of the leptonically decaying W_{lep} boson.

Their distributions and ratio $\frac{\text{Data}-\text{MC}_{\text{else}}}{\text{MC}_{W+\text{jets}}}$ are shown in the Figures 5.6 and 5.7. The correlation plots for these variables with $m_{\text{vis+met}}^{SH/HH}$ are shown in Section A.1. Their respective correlation factors are listed in Table 5.2 and 5.3. In contrast to the small correlation factors of the previous variables with $m_{\text{vis+met}}^{SH/HH}$, the transverse momenta of the $H \to b\bar{b}$

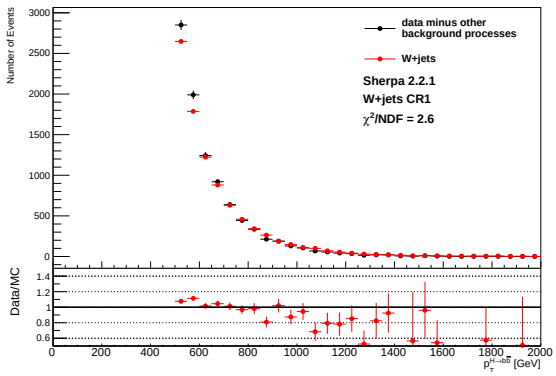
candidate $p_T^{H\to b\bar{b}}$ and the W_{had} candidate $p_T^{W_{\text{had}}}$ feature a significant correlation to the visible+met mass $m_{\text{vis+met}}^{SH/HH}$. This can be explained with the definition of $m_{\text{vis+met}}^{SH/HH}$ where the momenta p of the $H \to b\bar{b}$ candidate and W_{had} are included.



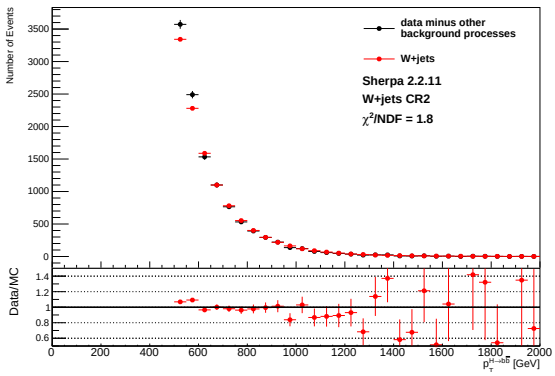
(a) $H \to b\bar{b}\ C_2$ (SHERPA 2.2.1)



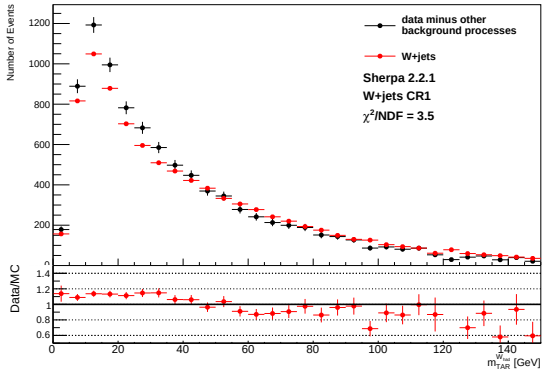
(b) $H \to b\bar{b}\ C_2$ (SHERPA 2.2.11)



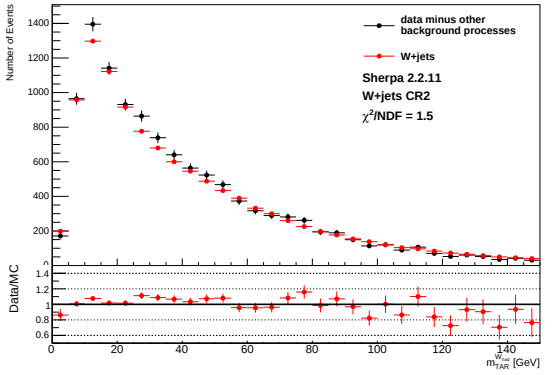
(c) $p_T^{H\to b\bar{b}}$ (SHERPA 2.2.1)



(d) $p_T^{H\to b\bar{b}}$ (SHERPA 2.2.11)

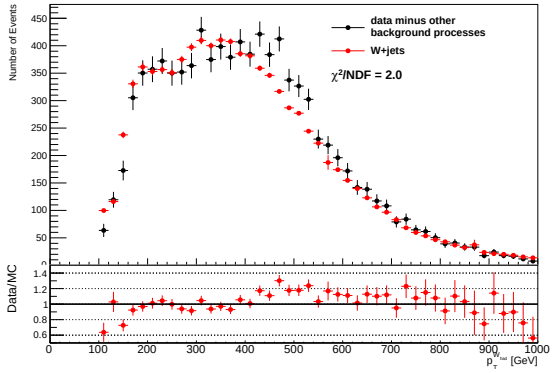


(e) $m_{\text{TAR}}^{\text{W}_{\text{had}}}$ (SHERPA 2.2.1)

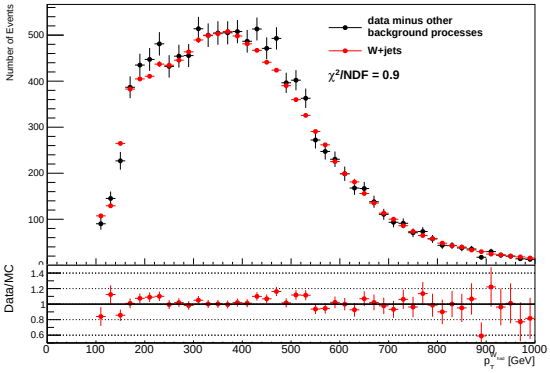


(f) $m_{\text{TAR}}^{\text{W}_{\text{had}}}$ (SHERPA 2.2.11)

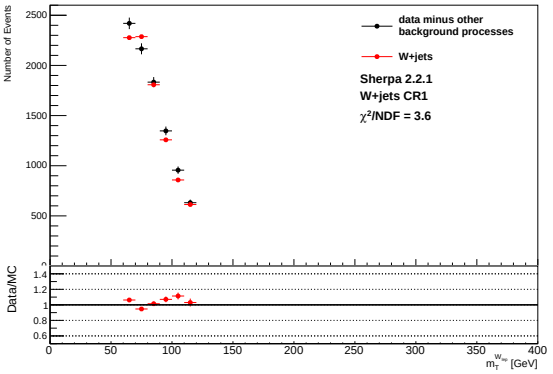
Figure 5.6.: The distributions of $H \to b\bar{b}\ C_2$, $p_T^{H\to b\bar{b}}$ and $m_{\text{TAR}}^{\text{W}_{\text{had}}}$ for SHERPA 2.2.1 (left side) and SHERPA 2.2.11 (right side) W+jets and data, respectively of which other background processes have been subtracted.



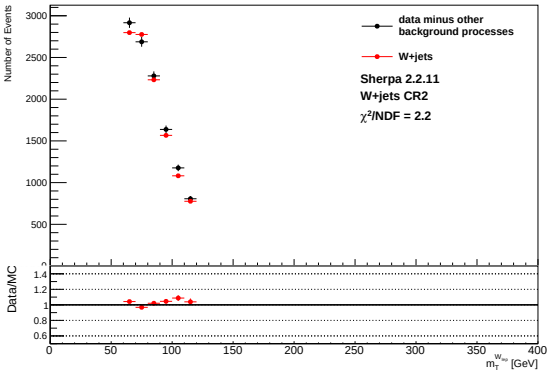
(a) $p_T^{W_{\text{had}}}$ (SHERPA 2.2.1)



(b) $p_T^{W_{\text{had}}}$ (SHERPA 2.2.11)



(c) $m_T^{W_{\text{lep}}}$ (SHERPA 2.2.1)



(d) $m_T^{W_{\text{lep}}}$ (SHERPA 2.2.11)

Figure 5.7.: The distributions of $p_T^{W_{\text{had}}}$ and $m_T^{W_{\text{lep}}}$ for SHERPA 2.2.1 (left side) and SHERPA 2.2.11 (right side) W+jets and data, respectively of which other background processes have been subtracted.

5.2.2. Reweighting the Background in the W+jets CRs

The ratio $\frac{\text{Data}-\text{MC}_{\text{else}}}{\text{MC}_{W+\text{jets}}}$ that has been obtained from the distributions in Figure 5.3, 5.4, 5.6 and 5.7 will be used as a weight on W+jets events for all the distributions to see if the modelling improves in other distributions in both W+jets CRs despite the five variables being basically uncorrelated.

The Tables 5.4 and 5.5 lists the χ^2/NDF values between the SHERPA 2.2.1 and SHERPA 2.2.11 data and weighted W+jets background. The χ^2/NDF values between the the nominal and weighted W+jets distributions from Table 5.1 are listed as well. As expected, if a distribution is weighted with its own ratio $\frac{\text{Data}-\text{MC}_{\text{else}}}{\text{MC}_{W+\text{jets}}}$, then the data and W+jets are in total agreement. This can be seen in the histograms for data and background in Section A.2 for each variable.

It can be observed that only the SHERPA 2.2.1 W+jets background of $m_{\text{vis+met}}^{SH/HH}$ is improved significantly by applying weights to the W+jets events. By using the weight $\frac{\text{Data}-\text{MC}_{\text{else}}}{\text{MC}_{W+\text{jets}}}$ from the W+jets distribution of $p_T^{H\to b\bar{b}}$ and $m_{\text{TAR}}^{\text{W}_{\text{had}}}$, the χ^2/NDF value improves from 6.0 to 3.4, 5.1 respectively. They are marked in light blue and violet in Table 5.4.

Similar results are seen for the SHERPA 2.2.11 W+jets background. Only the χ^2/NDF values regarding the W+jets background of $m_{\text{vis+met}}^{SH/HH}$ improves significantly. Here, the weight $\frac{\text{Data}-\text{MC}_{\text{else}}}{\text{MC}_{W+\text{jets}}}$ from the W+jets distribution of $p_T^{H\to b\bar{b}}$ improves the W+jets modelling. Furthermore, the weight from the W+jets distribution of $p_T^{W_{\text{had}}}$ improves the W+jets modelling regarding $m_{\text{vis+met}}^{SH/HH}$ as well. They are coloured in Table 5.5. The ratio $\frac{\text{Data}-\text{MC}_{\text{else}}}{\text{MC}_{W+\text{jets}}}$ from W+jets distribution of E_T^{miss} reduces the mismodelling of the W+jets background of $m_{\text{vis+met}}^{SH/HH}$ slightly.

Sherpa 2.2.1					
	$\Delta R(W_{\text{had}}, \ell)$	E_T^{miss}	lepton p_T	$m_{\text{TAR}}^{H\to b\bar{b}}$	$m_{\text{vis+met}}^{SH/HH}$
unweighted	**7.1**	**2.2**	**1.9**	**1.3**	**6.1**
weighted with					
$\Delta R(W_{\text{had}}, \ell)$	**0.0**	2.1	1.8	1.3	5.9
E_T^{miss}	6.9	**0.0**	1.9	1.4	5.8
$m_{\text{TAR}}^{H\to b\bar{b}}$	7.0	2.2	**0.0**	1.3	6.5
lepton p_T	7.0	2.3	1.9	**0.0**	6.1
$m_{\text{vis+met}}^{SH/HH}$	6.9	2.0	2.0	1.2	**0.0**
$H \to b\bar{b}$ C_2	—	—	—	—	6.1
$p_T^{H\to b\bar{b}}$	—	—	—	—	**3.4**
$m_{\text{TAR}}^{\text{W}_{\text{had}}}$	—	—	—	—	**5.1**
$p_T^{W_{\text{had}}}$	—	—	—	—	6.3
$m_T^{W_{\text{lep}}}$	—	—	—	—	6.2

Table 5.4.: The χ^2/NDF values between the weighted Sherpa 2.2.1 W+jets background and data and the χ^2/NDF values between the the nominal and weighted W+jets distributions from Table 5.1 are added additionally.

Sherpa 2.2.11					
	$\Delta R(W_{\text{had}}, \ell)$	E_T^{miss}	lepton p_T	$m_{\text{TAR}}^{H\to b\bar{b}}$	$m_{\text{vis+met}}^{SH/HH}$
unweighted	**2.2**	**2.5**	**1.3**	**1.1**	**4.0**
weighted with					
$\Delta R(W_{\text{had}}, \ell)$	**0.0**	2.5	1.3	1.2	4.0
E_T^{miss}	2.2	**0.0**	1.3	1.1	3.6
$m_{\text{TAR}}^{H\to b\bar{b}}$	2.2	2.5	**0.0**	1.1	3.8
lepton p_T	2.3	2.5	1.3	**0.0**	3.9
$m_{\text{vis+met}}^{SH/HH}$	2.2	2.2	1.3	1.1	**0.0**
$H \to b\bar{b}$ C_2	—	—	—	—	4.1
$p_T^{H\to b\bar{b}}$	—	—	—	—	**2.7**
$m_{\text{TAR}}^{\text{W}_{\text{had}}}$	—	—	—	—	3.7
$p_T^{W_{\text{had}}}$	—	—	—	—	**3.4**
$m_T^{W_{\text{lep}}}$	—	—	—	—	4.0

Table 5.5.: The χ^2/NDF values between the weighted Sherpa 2.2.11 W+jets background and data and the χ^2/NDF values between the the nominal and weighted W+jets distributions from Table 5.1 are added additionally.

The Figures 5.8 and 5.9 show the nominal and reweighted SHERPA 2.2.1, SHERPA 2.2.11 W+jets distributions for $m_{\text{vis+met}}^{SH/HH}$, respectively. Just as the χ^2/NDF values indicate, the best improvement follows from the applied weight $\frac{\text{Data}-\text{MC}_{\text{else}}}{\text{MC}_{W+\text{jets}}}$ from the SHERPA 2.2.1 and 2.2.11 W+jets distribution of $p_T^{H\to b\bar{b}}$. As expected, the SHERPA 2.2.1 and 2.2.11 W+jets background modelling improves significantly for $m_{\text{vis+met}}^{SH/HH}$ when the W+jets events are weighted with the ratio $\frac{\text{Data}-\text{MC}_{\text{else}}}{\text{MC}_{W+\text{jets}}}$ from the W+jets distribution of $p_T^{H\to b\bar{b}}$. The underestimation of data is smaller for the region around the peak. The same can be observed for the region $m_{\text{vis+met}}^{SH/HH} > 2000$ GeV where the modelling in the W+jets CRs exhibits a general better agreement with data: Using $p_T^{H\to b\bar{b}}$ as a weight for the MC simulation, the overestimation of background is still present but the simulation is significantly closer to the data points. The weights coming from the other variables that improves the SHERPA 2.2.1 and 2.2.11 W+jets background of $m_{\text{vis+met}}^{SH/HH}$ are also shown.

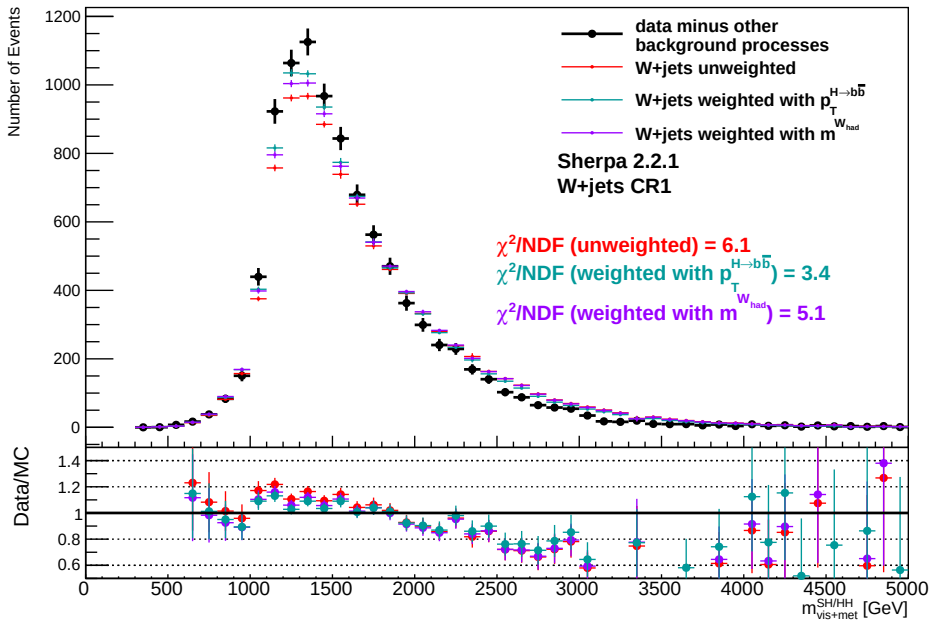


Figure 5.8.: The nominal and reweighted $m_{\text{vis+met}}^{SH/HH}$ histogram in the W+jets CR1 (SHERPA 2.2.1). The ratio $\frac{\text{Data}-\text{MC}_{\text{else}}}{\text{MC}_{W+\text{jets}}}$ from $p_T^{H\to b\bar{b}}$ and $m_{\text{TAR}}^{\text{W}_{\text{had}}}$ distributions is used as the weight for the W+jets background simulation.

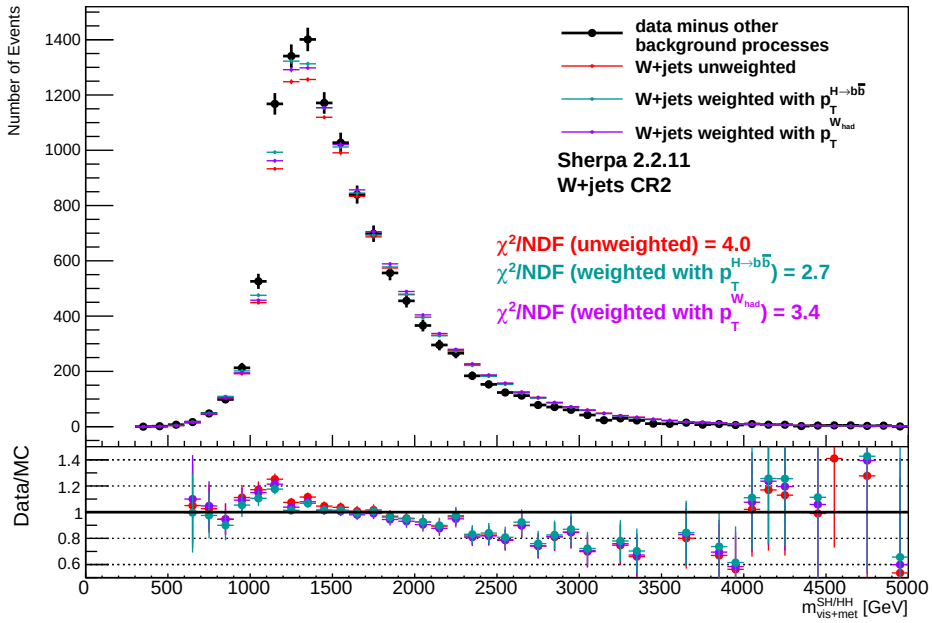


Figure 5.9.: The nominal and reweighted $m_{\text{vis+met}}^{SH/HH}$ histogram in the W+jets CR2 (SHERPA 2.2.11). The ratio $\frac{\text{Data}-\text{MC}_{\text{else}}}{\text{MC}_{W+\text{jets}}}$ from $p_T^{H\to b\bar{b}}$ and $m_{\text{TAR}}^{\text{W}_{\text{had}}}$ distributions is used as the weight for the W+jets background simulation.

6. Modelling in the Validation Regions

In Chapter 5, W+jets background in both W+jets CRs using the SHERPA versions 2.2.1 and 2.2.11 has been investigated. Reweighting the W+jets events marginally improved the W+jets modelling. Here, the reweighting is investigated in the VR. The weights obtained in the W+jets CR1 and W+jets CR2 for SHERPA 2.2.1 and SHERPA 2.2.11 W+jets samples, respectively, are used because these regions have the largest W+jets contribution.

The focus is placed on $m_{\text{vis+met}}^{SH/HH}$ distribution in the VR and VR_{spec} (both introduced in Section 4.4.4) as it is the most important variable in this analysis. Here, it is important to mention, that the SHERPA 2.2.11 background modelling is not considered due to the small number of events passing the VR SH and especially HH selection for the previously investigated main variables ($\Delta R(W_{\text{had}}, \ell)$, E_T^{miss}, $m_{\text{TAR}}^{H\to b\bar{b}}$, lepton p_T and $m_{\text{vis+met}}^{SH/HH}$). Thus, SHERPA 2.2.11 is only considered for VR_{spec}.

6.1. Distribution of $m_{\text{vis+met}}^{SH/HH}$ in the VR

The background modelling for $m_{\text{vis+met}}^{SH/HH}$ in the VR is shown in the Figures 6.1 and 6.2. It is distinguished between the SH and HH selection. As a reminder, the VR cuts are:

- 1 b-tagged TAR jet in the event,
- $H \to b\bar{b}$ candidate is single b-tagged and
- $H \to b\bar{b}$ candidate fails the 80 % m_H mass window,
- *SH selection:* $H \to b\bar{b}$ candidate passes 80 % C_2 window and
- *HH selection:* event candidate passes the SH selection criteria and passes 80 % $\Delta R(W_{\text{had}}, \ell)$ window.

Compared to the background modelling in the W+jets CR1 and W+jets CR2 in the Figures 5.2(e) and 5.2(f), the first thing to mention are the larger statistical uncertainties in the VR. This is the result of less events selected in the VR. For SHERPA 2.2.11 W+jets, the window cut on $H \to b\bar{b}$ C_2 has been rederived and removes a large amount of background compared to the previous definition. Due to the extra cut for the HH selection in the VR, even less events are saved than for the SH selection which results in even larger statistical uncertainties. Secondly, for all the five previously investigated variables, the other background processes have a much larger contribution in the VR than in the W+jets CRs, such that for SHERPA 2.2.11 W+jets, it is not the leading background process.

SH Selection

The background modelling for $m_{\text{vis+met}}^{SH/HH}$ in the VR (SH selection) is shown in Figure 6.1.

Similar to the modelling in the W+jets CR1 and W+jets CR2, the background (Sherpa 2.2.1 in Figure 6.1) tends to underestimate the data around the peak (900 GeV to 2100 GeV) of the distribution.

The χ^2/NDF value (Equation 5.1) has been calculated here as well in order to compare W+jets distribution and data. For Sherpa 2.2.1, the χ^2/NDF value regarding $m_{\text{vis+met}}^{SH/HH}$ amounts to

$$\chi^2/\text{NDF} = 3.9. \tag{6.1}$$

With a value of χ^2/NDF= 6.1 for the Sherpa 2.2.1 W+jets background in the W+jets CR1 for $m_{\text{vis+met}}^{SH/HH}$, the improvement in the VR (SH selection) using Sherpa 2.2.1 for the W+jets background is significantly better.

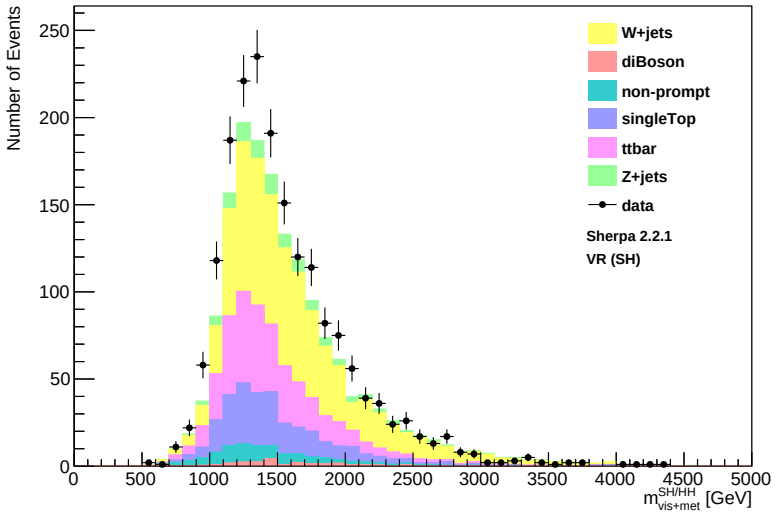


Figure 6.1.: The stacked histogram of $m_{\text{vis+met}}^{SH/HH}$ in the VR (SH selection, Sherpa 2.2.1).

HH Selection

The background modelling for $m_{\text{vis+met}}^{SH/HH}$ in the VR (HH selection) is shown in Figure 6.2.The background modelling for $m_{\text{vis+met}}^{SH/HH}$ (Sherpa 2.2.1) exhibits a better agreement with data for the HH selection compared to the modelling for the SH selection.

This is confirmed by the χ^2/NDF value between MC predicted W+jets distribution and data for Sherpa 2.2.1:

$$\chi^2/\text{NDF} = 3.7. \tag{6.2}$$

Similar to the χ^2 value of 3.9 for the Sherpa W+jets background in the VR (SH selection), the χ^2 value for the Sherpa 2.2.1 W+jets background in the VR (HH selection) is much smaller than $\chi^2 = 6.1$ for the Sherpa 2.2.1 W+jets background in the W+jets CR1 for $m_{\text{vis+met}}^{SH/HH}$.

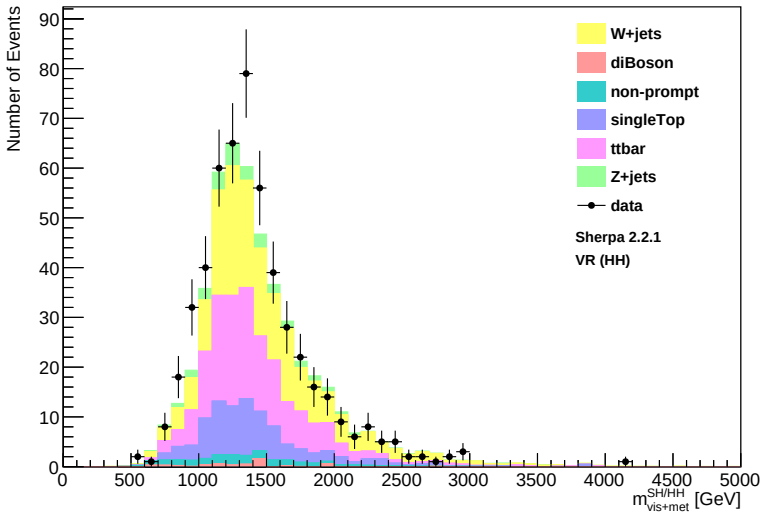


Figure 6.2.: The stacked histograms of $m_{\text{vis+met}}^{SH/HH}$ in the VR (HH selection, Sherpa 2.2.1).

6.2. Reweighting the Background in the VR

The weights $\frac{\text{Data}-\text{MC}_{\text{else}}}{\text{MC}_{W+\text{jets}}}$ that are applied to the Sherpa 2.2.1 W+jets background in the VR (SH and HH selection) have been derived in the W+jets CR1. The focus will be on the more promising variables in the VR to reweight the Sherpa 2.2.1 W+jets distribution of $m_{\text{vis+met}}^{SH/HH}$: $\Delta R(W_{\text{had}}, \ell)$, E_T^{miss}, $m_{\text{TAR}}^{H\to b\bar{b}}$, lepton p_T, $m_{\text{vis+met}}^{SH/HH}$, $p_T^{H\to b\bar{b}}$, $m_{\text{TAR}}^{\text{W}_{\text{had}}}$ and $p_T^{W_{\text{had}}}$.

For the χ^2/NDF values in Table 6.1 for $m_{\text{vis+met}}^{SH/HH}$, the Sherpa 2.2.1 W+jets modelling improves if the weight $\frac{\text{Data}-\text{MC}_{\text{else}}}{\text{MC}_{W+\text{jets}}}$ obtained from the $m_{\text{vis+met}}^{SH/HH}$ distribution in the W+jets

CR1 (SHERPA 2.2.1) is applied for both the SH and HH selection. The improvements are highlighted in light blue. To compare the unweighted and weighted W+jets distributions, they have been put together in Figure 6.3 and 6.4. These plots visualize the improvement of the W+jets background modelling for $m_{\text{vis+met}}^{SH/HH}$ after weighting the W+jets events. In both cases, the improvement of the underestimation of data around the peaks is clearly visible. The overestimation of data for higher $m_{\text{vis+met}}^{SH/HH}$ values is slightly improved but the disagreement between W+jets distribution and data is still present.

Reweighting based on the W+jets distribution of $p_T^{H\to b\bar{b}}$ also improves the W+jets agreement in the VR (SH selection) for $m_{\text{vis+met}}^{SH/HH}$ significantly (highlighted in violet in Table 6.1) and is shown in Figure 6.3. The weights coming from the W+jets distribution of $p_T^{H\to b\bar{b}}$ does not improve the modelling in the VR (HH selection) significantly for $m_{\text{vis+met}}^{SH/HH}$. The χ^2/NDF value only improves from 3.7 to 3.5.

The background modelling of $m_{\text{vis+met}}^{SH/HH}$ with the W+jets events weighted with other variables is explicitly shown in Section B.1.

VR | SHERPA 2.2.1	$\boldsymbol{\chi^2}$**/NDF**	
	$m_{\text{vis+met}}^{SH/HH}$	
	SH selection	HH selection
unweighted	**3.9**	**3.7**
weighted with		
$\Delta R(W_{\text{had}}, \ell)$	3.9	3.6
E_T^{miss}	3.8	3.6
$m_{\text{TAR}}^{H\to b\bar{b}}$	3.9	3.6
lepton p_T	3.8	3.8
$m_{\text{vis+met}}^{SH/HH}$	**3.1**	**2.8**
$p_T^{H\to b\bar{b}}$	**3.4**	3.5
$m_{\text{TAR}}^{\text{W}_{\text{had}}}$	3.8	3.6
$p_T^{W_{\text{had}}}$	3.9	3.7

Table 6.1.: The χ^2/NDF values for the nominal and weighted SHERPA 2.2.1 W+jets background in the VR (SH and HH selection) for $m_{\text{vis+met}}^{SH/HH}$. The numbers highlighted show the largest improvements.

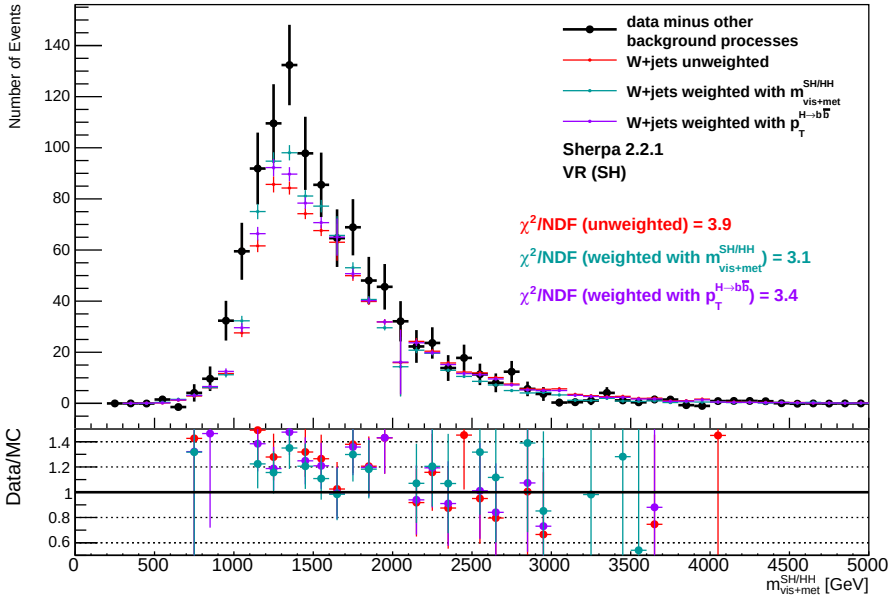


Figure 6.3.: The unweighted and weighted SHERPA 2.2.1 W+jets distribution for $m^{SH/HH}_{\text{vis+met}}$ in the VR (SH selection).

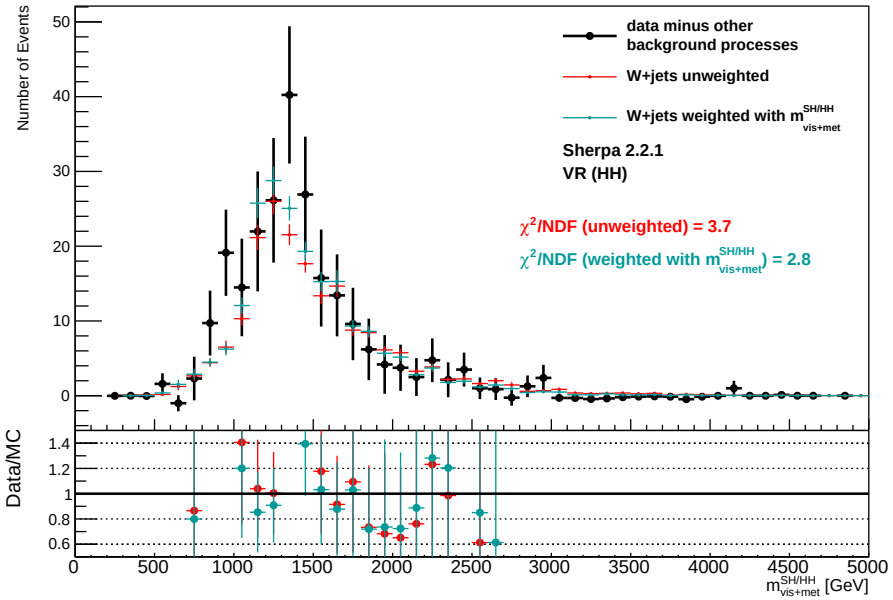
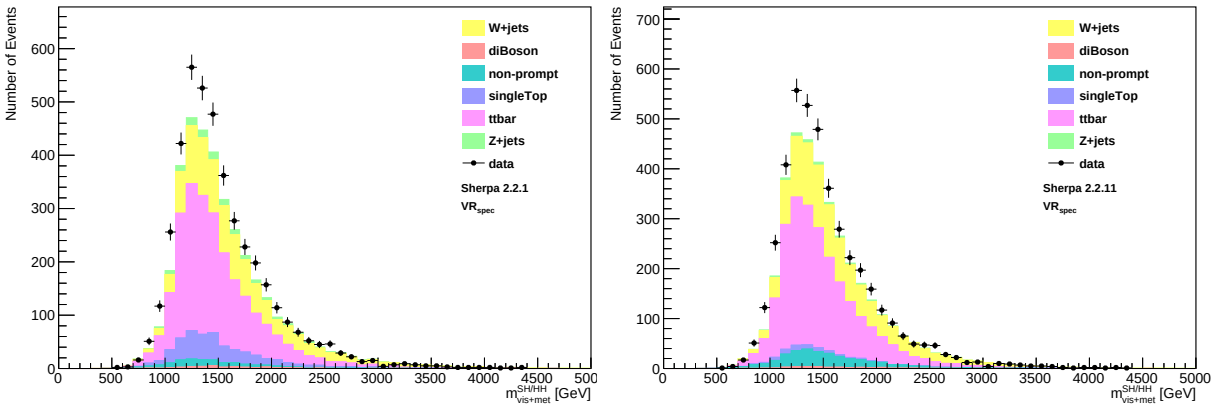


Figure 6.4.: The unweighted and weighted SHERPA 2.2.1 W+jets distribution for $m^{SH/HH}_{\text{vis+met}}$ in the VR (HH selection).

6.3. VR$_{\text{spec}}$

In Section 4.4.4, the special VR denoted as VR$_{\text{spec}}$ was introduced to reduce statistical fluctuations especially after rederiving the $H \to b\bar{b}$ C_2 window cut, and thus obtain a better handle on the modelling. There is no distinction between SH and HH selection in this case. The distribution in VR$_{\text{spec}}$ for $m_{\text{vis+met}}^{SH/HH}$ using both SHERPA versions is shown in Figure 6.5. The underestimation around the peak is clearly observable in both plots. In contrast to the background modelling in the W+jets CRs, the tail of the distribution is underestimated in the VR$_{\text{spec}}$.



(a) SHERPA 2.2.1. ***(b)*** SHERPA 2.2.11.

Figure 6.5.: Comparison between expected backgrounds (filled histograms) and data (black points) for $m_{\text{vis+met}}^{SH/HH}$ in the VR$_{\text{spec}}$.

The χ^2/NDF values for the unweighted and weighted comparisons between W+jets simulated with SHERPA 2.2.1 and SHERPA 2.2.11 and data are listed in Table 6.2. The SHERPA 2.2.1 W+jets modelling in the VR$_{\text{spec}}$ for $m_{\text{vis+met}}^{SH/HH}$ improves with the weights taken from $m_{\text{vis+met}}^{SH/HH}$, $p_T^{H\to b\bar{b}}$ and $\Delta R(W_{\text{had}}, \ell)$ in the W+jets CR1. For the SHERPA 2.2.11 W+jets modelling, the weights from $p_T^{H\to b\bar{b}}$ and $p_T^{W_{\text{had}}}$ improve the modelling slightly. All improvements are coloured in Table 6.2 and Figure 6.6 and 6.7 show the unweighted and weighted SHERPA 2.2.1 and SHERPA 2.2.11 W+jets distribution and data in the VR$_{\text{spec}}$ for $m_{\text{vis+met}}^{SH/HH}$. Even though the χ^2/NDF values improve for the previously mentioned variables whose W+jets ratio $\frac{\text{Data}-\text{MC}_{\text{else}}}{\text{MC}_{W+\text{jets}}}$ have been applied to the SHERPA 2.2.1 and 2.2.11 W+jets distribution of $m_{\text{vis+met}}^{SH/HH}$, the underestimation of data for the tail of the W+jets distribution and especially around the peak is still present.

The corresponding stacked histograms with the weighted W+jets background is shown in Section B.2.

VR_{spec} | SHERPA 2.2.1	χ^2/NDF $m_{\text{vis+met}}^{SH/HH}$
unweighted	**4.7**
weighted with	
$\Delta R(W_{\text{had}}, \ell)$	**4.3**
E_T^{miss}	4.5
$m_{\text{TAR}}^{H\to b\bar{b}}$	4.8
lepton p_T	4.6
$m_{\text{vis+met}}^{SH/HH}$	**3.8**
$p_T^{H\to b\bar{b}}$	**4.1**
$m_{\text{TAR}}^{W_{\text{had}}}$	4.5
$p_T^{W_{\text{had}}}$	4.8

(a)

VR_{spec} | SHERPA 2.2.11	χ^2/NDF $m_{\text{vis+met}}^{SH/HH}$
unweighted	**4.9**
weighted with	
$\Delta R(W_{\text{had}}, \ell)$	4.9
E_T^{miss}	4.8
$m_{\text{TAR}}^{H\to b\bar{b}}$	4.8
lepton p_T	4.9
$m_{\text{vis+met}}^{SH/HH}$	4.9
$p_T^{H\to b\bar{b}}$	**4.6**
$m_{\text{TAR}}^{W_{\text{had}}}$	4.9
$p_T^{W_{\text{had}}}$	**4.5**

(b)

Table 6.2.: The χ^2/NDF values for the nominal and weighted (a) SHERPA 2.2.1 and (b) SHERPA 2.2.11 W+jets background in the VR_{spec} for $m_{\text{vis+met}}^{SH/HH}$.

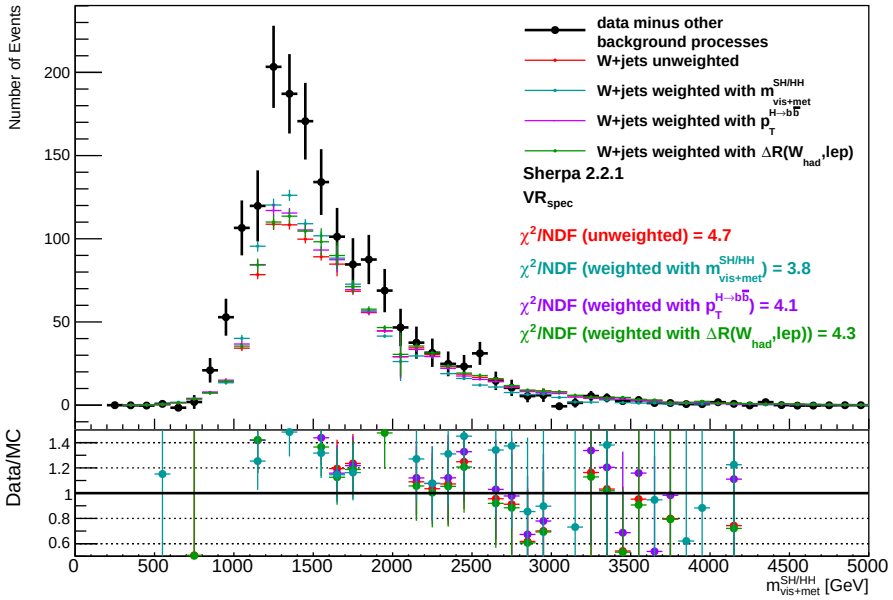


Figure 6.6.: The unweighted and weighted SHERPA 2.2.1 W+jets distribution for $m_{\text{vis+met}}^{SH/HH}$ in the VR_{spec}.

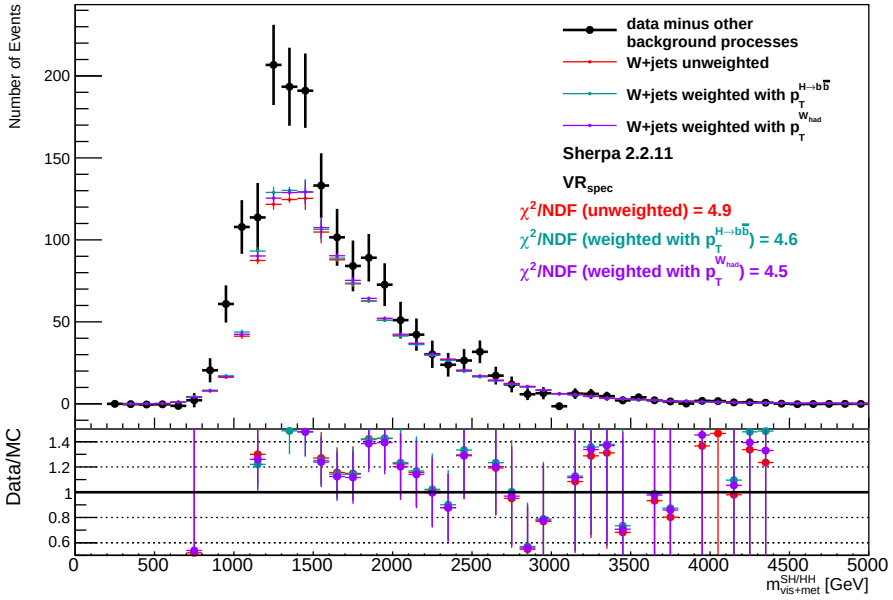


Figure 6.7.: The unweighted and weighted SHERPA 2.2.11 W+jets distribution for $m_{\mathrm{vis+met}}^{SH/HH}$ in the $\mathrm{VR}_{\mathrm{spec}}$.

Ultimately, the SHERPA 2.2.1 and SHERPA 2.2.11 W+jets background modelling for $m_{\mathrm{vis+met}}^{SH/HH}$ in the VR and $\mathrm{VR}_{\mathrm{spec}}$ agrees better with data than in the W+jets CRs. After reweighting the W+jets events, the W+jets distribution for $m_{\mathrm{vis+met}}^{SH/HH}$ improves slightly, especially with the weights $\frac{\mathrm{Data}-\mathrm{MC}_{\mathrm{else}}}{\mathrm{MC}_{W+\mathrm{jets}}}$ derived from the $p_T^{H\to b\bar{b}}$ distribution in the respective W+jets CR.

7. Conclusion and Outlook

In this thesis, the W+jets background modelling in the W+jets CRs and VR for five different variables was investigated. For the W+jets background, the sample simulated by Sherpa 2.2.1 was used and later upgraded to the Sherpa 2.2.11 sample.

Firstly, it can be said that in the W+jets CR the Sherpa 2.2.11 W+jets events yield a better agreement with data than the W+jets background from Sherpa 2.2.1. The χ^2/NDF values between data and the W+jets events are smaller, especially for $m_{\text{vis+met}}^{SH/HH}$ where the W+jets background underestimates and overestimates the data in the peak and in the tail, respectively. Due to the importance of $m_{\text{vis+met}}^{SH/HH}$ in the entire analysis, an improvement of the W+jets modelling in this variable is favourable. By reweighting the W+jets events in the W+jets CR with the weight $\frac{\text{Data}-\text{MC}_{\text{else}}}{\text{MC}_{W+\text{jets}}}$ obtained from the $p_T^{H\to b\bar{b}}$ distribution, the modelling improves significantly for both Sherpa samples due to the non-vanishing correlation between $m_{\text{vis+met}}^{SH/HH}$ and $p_T^{H\to b\bar{b}}$. In addition, the applied weight $\frac{\text{Data}-\text{MC}_{\text{else}}}{\text{MC}_{W+\text{jets}}}$ from the $m_{\text{TAR}}^{\text{W}_{\text{had}}}$ distribution (Sherpa 2.2.1) and $p_T^{W_{\text{had}}}$ (Sherpa 2.2.11) reduces the overestimation and underestimation of data as well, but not as much as the weights derived from $p_T^{H\to b\bar{b}}$. Using the ratio $\frac{\text{Data}-\text{MC}_{\text{else}}}{\text{MC}_{W+\text{jets}}}$ from other variables, such as $\Delta R(W_{\text{had}}, \ell)$, E_T^{miss}, $m_{\text{TAR}}^{H\to b\bar{b}}$, lepton p_T, $H \to b\bar{b}$ C_2 and $m_T^{W_{\text{lep}}}$, as a weighting factor, the W+jets modelling for $m_{\text{vis+met}}^{SH/HH}$ does not improve.

Similar observations can be made in the VR for $m_{\text{vis+met}}^{SH/HH}$. A disagreement between the predicted W+jets background and data is clearly visible. The modelling of the W+jets background improves using the new Sherpa sample. Applying the weight $\frac{\text{Data}-\text{MC}_{\text{else}}}{\text{MC}_{W+\text{jets}}}$ that has been obtained from $m_{\text{vis+met}}^{SH/HH}$ in the W+jets CR yields the best improvement. This is valid for both SH and HH selections. Similar to the W+jets CRs, the weight coming from the W+jets distribution of $p_T^{H\to b\bar{b}}$ improves the modelling in the VR but only slightly. The same is true for the Sherpa 2.2.1 and Sherpa 2.2.11 W+jets distributions of $m_{\text{vis+met}}^{SH/HH}$ in the VR_{spec}.

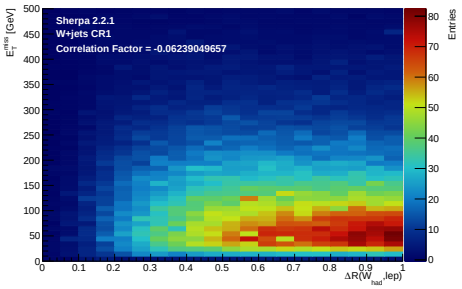
Therefore, reweighting the W+jets events in the distribution of $m_{\text{vis+met}}^{SH/HH}$ improves the W+jets modelling. The weights from variables that are correlated to $m_{\text{vis+met}}^{SH/HH}$ improves

the W+jets background modelling of $m_{\text{vis+met}}^{SH/HH}$ in the W+jets CR and in the VR and therefore, should be considered to minimise the disagreement between data and MC simulated W+jets events.

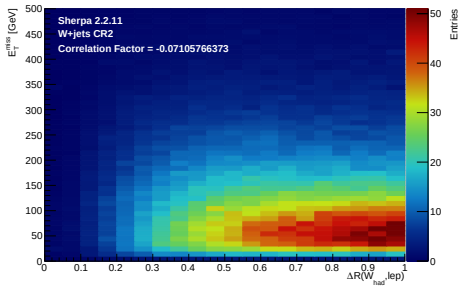
For further steps in the analysis, SHERPA 2.2.11 should be used for the W+jets background modelling due to the better agreement between data and MC simulated W+jets background. It might also improve the background modelling in the CRs of the other two background processes, $t\bar{t}$ and especially QCD. Furthermore, the weight $\frac{\text{Data}-\text{MC}_{\text{else}}}{\text{MC}_{W+\text{jets}}}$ from other, in this thesis not evaluated variables should be considered to improve the W+jets background modelling in the distributions of variables and especially $m_{\text{vis+met}}^{SH/HH}$. A better background modelling and thus, a better understanding of the background would help in the search of non-resonant Higgs pair production and would strengthen the analysis sensitivity to corresponding scenarios beyond the SM.

A. W+jets CRs

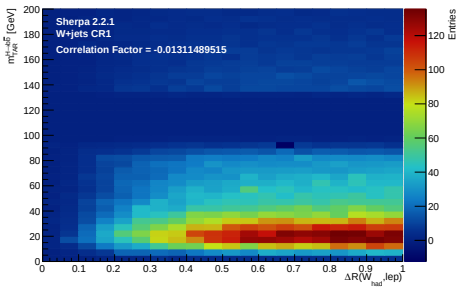
A.1. Correlation Plots in the W+jets CRs



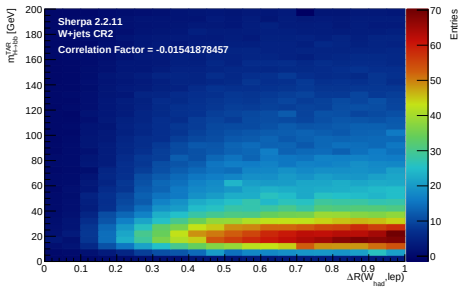
(a) Correlation between $\Delta R(W_{\text{had}}, \ell)$ and E_T^{miss} (SHERPA 2.2.1)



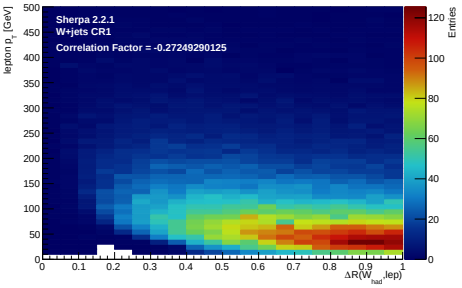
(b) Correlation between $\Delta R(W_{\text{had}}, \ell)$ and E_T^{miss} (SHERPA 2.2.11)



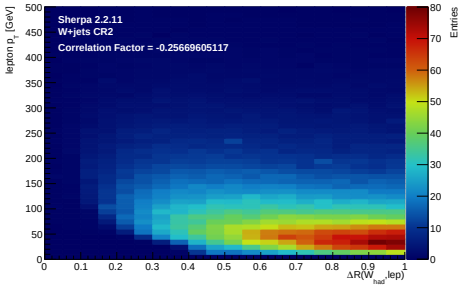
(c) Correlation between $\Delta R(W_{\text{had}}, \ell)$ and $m_{\text{TAR}}^{H\to b\bar{b}}$ (SHERPA 2.2.1)



(d) Correlation between $\Delta R(W_{\text{had}}, \ell)$ and $m_{\text{TAR}}^{H\to b\bar{b}}$ (SHERPA 2.2.11)

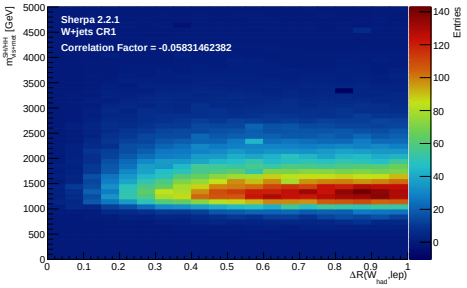


(e) Correlation between $\Delta R(W_{\text{had}}, \ell)$ and lepton p_T (SHERPA 2.2.1)

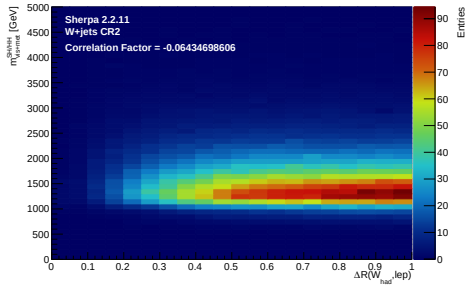


(f) Correlation between $\Delta R(W_{\text{had}}, \ell)$ and lepton p_T (SHERPA 2.2.11)

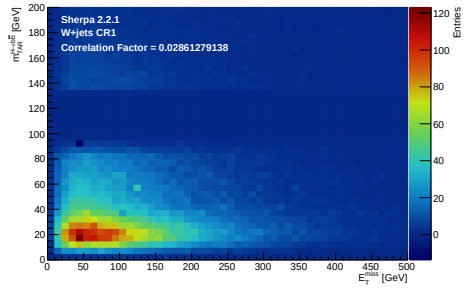
Figure A.1.: The correlation plots with the correlation factors between different variables in the W+jets CR1 (left plots) and W+jets CR2 (right plots).



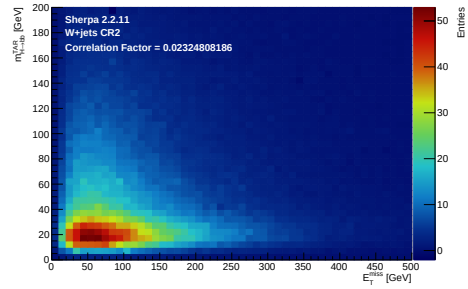
(a) Correlation between $\Delta R(W_{\text{had}}, \ell)$ and $m_{\text{vis+met}}^{SH/HH}$ (Sherpa 2.2.1)



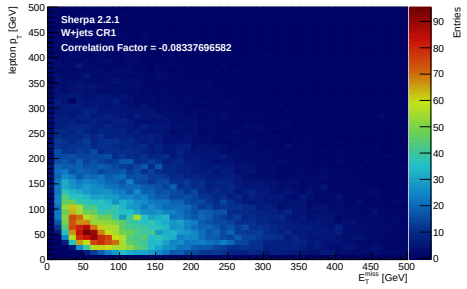
(b) Correlation between $\Delta R(W_{\text{had}}, \ell)$ and $m_{\text{vis+met}}^{SH/HH}$ (Sherpa 2.2.11)



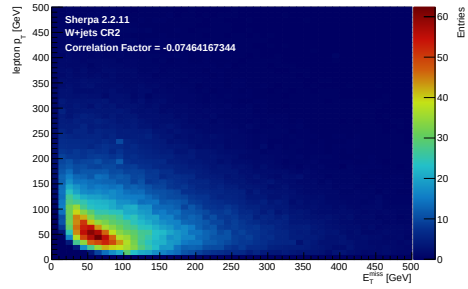
(c) Correlation between E_T^{miss} and $m_{\text{TAR}}^{H\to b\bar{b}}$ (Sherpa 2.2.1)



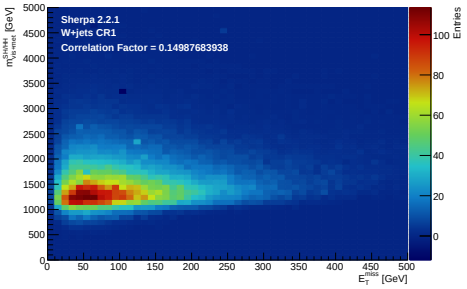
(d) Correlation between E_T^{miss} and $m_{\text{TAR}}^{H\to b\bar{b}}$ (Sherpa 2.2.11)



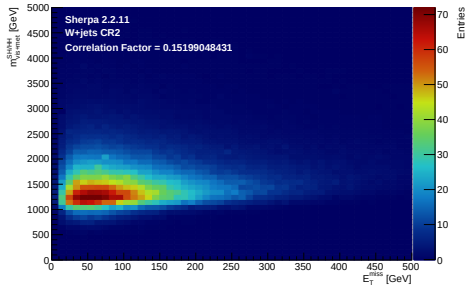
(e) Correlation between E_T^{miss} and lepton p_T (Sherpa 2.2.1)



(f) Correlation between E_T^{miss} and lepton p_T (Sherpa 2.2.11)

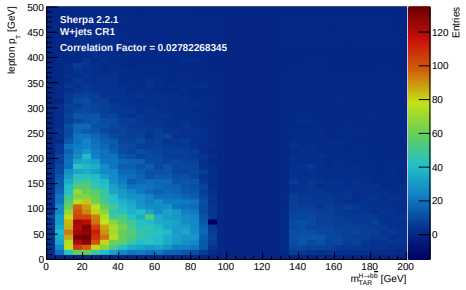


(g) Correlation between E_T^{miss} and $m_{\text{vis+met}}^{SH/HH}$ (Sherpa 2.2.1)

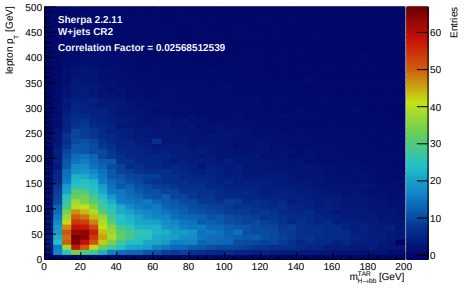


(h) Correlation between E_T^{miss} and $m_{\text{vis+met}}^{SH/HH}$ (Sherpa 2.2.11)

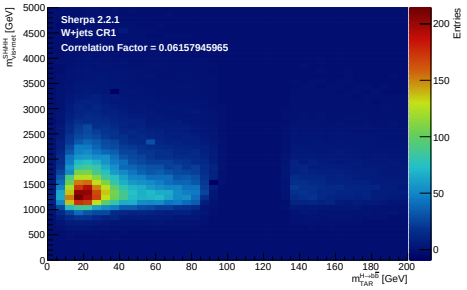
Figure A.2.: The correlation plots with the correlation factors between different variables in the W+jets CR1 (left plots) and W+jets CR2 (right plots).



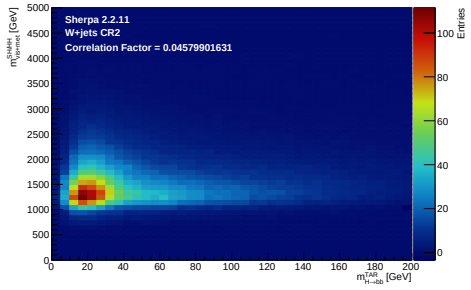
(a) Correlation between $m^{H\to b\bar{b}}_{\text{TAR}}$ and lepton p_T (SHERPA 2.2.1)



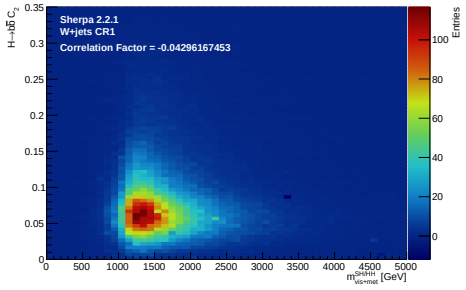
(b) Correlation between $m^{H\to b\bar{b}}_{\text{TAR}}$ and lepton p_T (SHERPA 2.2.11)



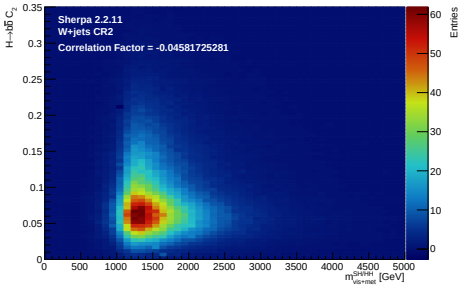
(c) Correlation between $m^{H\to b\bar{b}}_{\text{TAR}}$ and $m^{SH/HH}_{\text{vis+met}}$ (SHERPA 2.2.1)



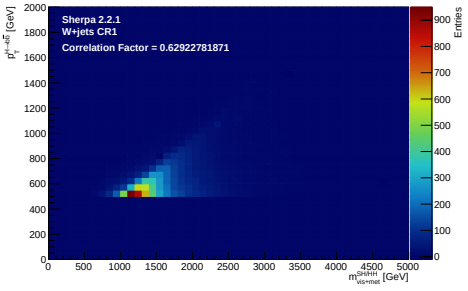
(d) Correlation between $m^{H\to b\bar{b}}_{\text{TAR}}$ and $m^{SH/HH}_{\text{vis+met}}$ (SHERPA 2.2.11)



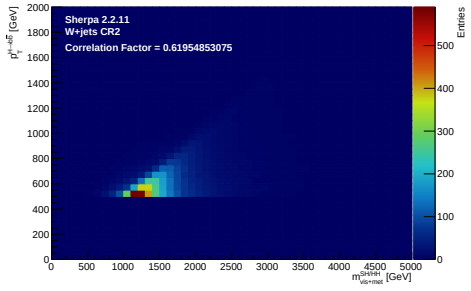
(e) Correlation between $m^{SH/HH}_{\text{vis+met}}$ and $H \to b\bar{b}$ C_2 (SHERPA 2.2.1)



(f) Correlation between $m^{SH/HH}_{\text{vis+met}}$ and $H \to b\bar{b}$ C_2 (SHERPA 2.2.11)

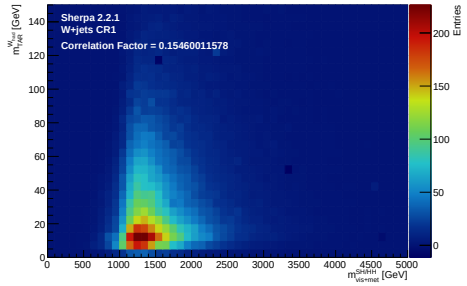


(g) Correlation between $m^{SH/HH}_{\text{vis+met}}$ and $p_T^{H\to b\bar{b}}$ (SHERPA 2.2.1)

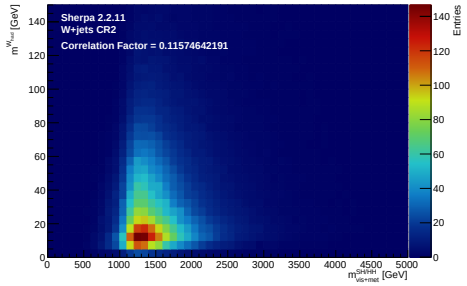


(h) Correlation between $m^{SH/HH}_{\text{vis+met}}$ and $p_T^{H\to b\bar{b}}$ (SHERPA 2.2.11)

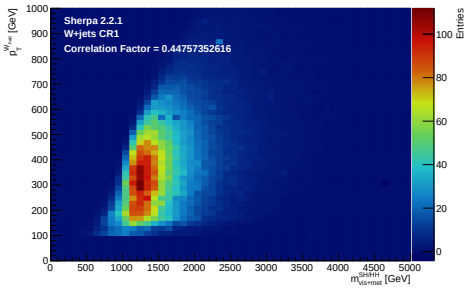
Figure A.3.: The correlation plots with the correlation factors between different variables in the W+jets CR1 (left plots) and W+jets CR2 (right plots).



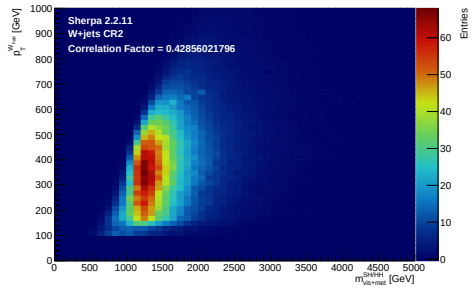
(a) Correlation between $m_{\text{vis+met}}^{SH/HH}$ and $m_{\text{TAR}}^{\text{W}_{\text{had}}}$ (Sherpa 2.2.1)



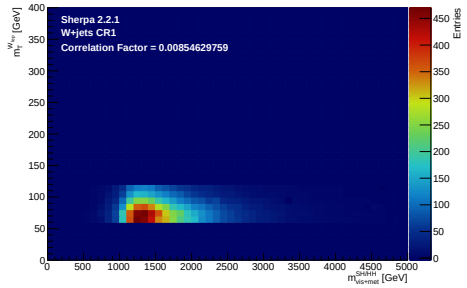
(b) Correlation between $m_{\text{vis+met}}^{SH/HH}$ and $m_{\text{TAR}}^{\text{W}_{\text{had}}}$ (Sherpa 2.2.11)



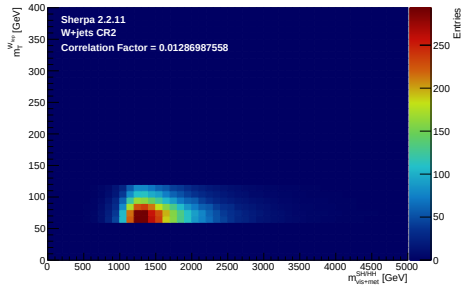
(c) Correlation between $m_{\text{vis+met}}^{SH/HH}$ and $p_T^{W_{\text{had}}}$ (Sherpa 2.2.1)



(d) Correlation between $m_{\text{vis+met}}^{SH/HH}$ and $p_T^{W_{\text{had}}}$ (Sherpa 2.2.11)



(e) Correlation between $m_{\text{vis+met}}^{SH/HH}$ and $m_T^{W_{\text{lep}}}$ (Sherpa 2.2.1)



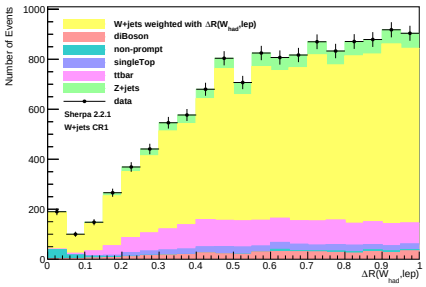
(f) Correlation between $m_{\text{vis+met}}^{SH/HH}$ and $m_T^{W_{\text{lep}}}$ (Sherpa 2.2.11)

Figure A.4.: The correlation plots and the correlation factors between different variables in the W+jets CR1 (left plots) and W+jets CR2 (right plots).

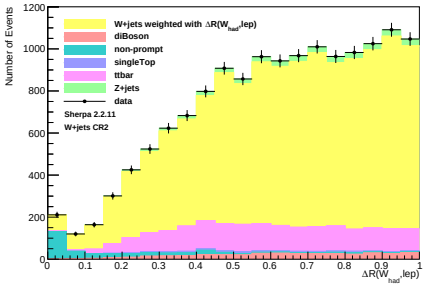
A.2. Weighted W+jets Events in the W+jets CRs

Here, "Weighted with [variable]" means that the W+jets events are weighted with $\frac{\text{Data}-\text{MC}_{\text{else}}}{\text{MC}_{W+\text{jets}}}$ obtained from the distribution of the variable in the respective W+jets CR.

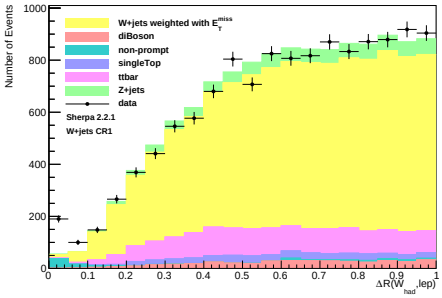
A.2.1. $\Delta R(W_{\text{had}}, \ell)$



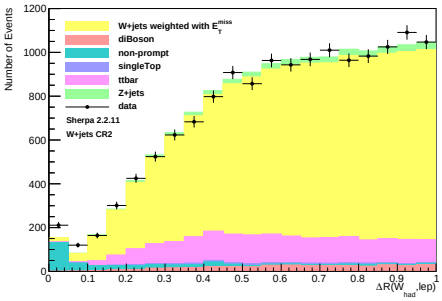
(a) Weighted with $\Delta R(W_{\text{had}}, \ell)$ (Sherpa 2.2.1)



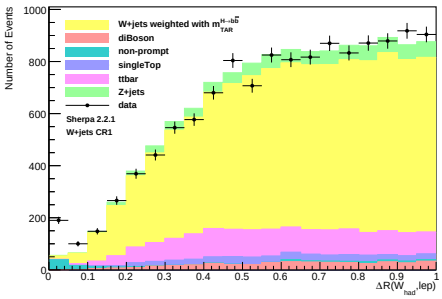
(b) Weighted with $\Delta R(W_{\text{had}}, \ell)$ (Sherpa 2.2.11)



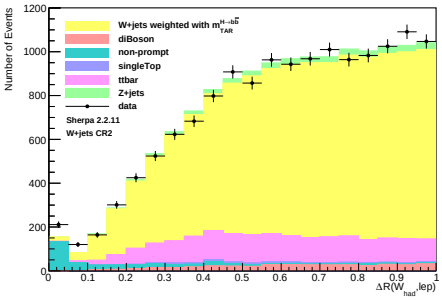
(c) Weighted with E_T^{miss} (Sherpa 2.2.1)



(d) Weighted with E_T^{miss} (Sherpa 2.2.11)

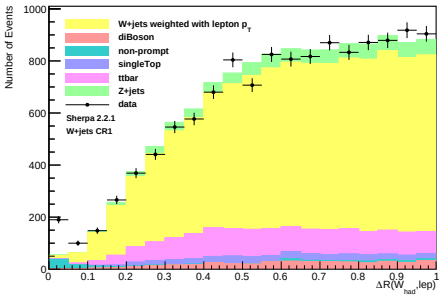


(e) Weighted with $m_{\text{TAR}}^{H\to b\bar{b}}$ (Sherpa 2.2.1)

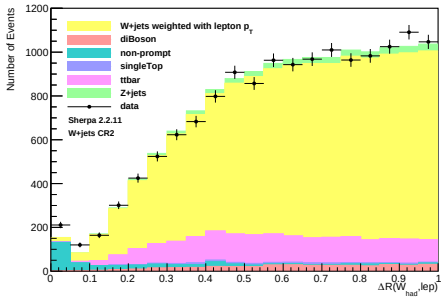


(f) Weighted with $m_{\text{TAR}}^{H\to b\bar{b}}$ (Sherpa 2.2.11)

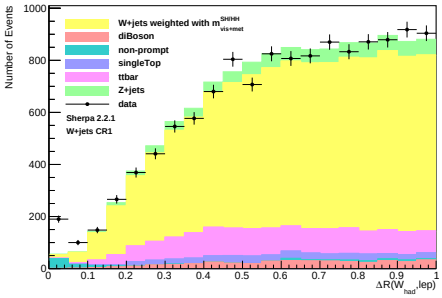
Figure A.5.: The histograms with the stacked background processes for $\Delta R(W_{\text{had}}, \ell)$ in the W+jets CR1 (left plots) and W+jets CR2 (right plots).



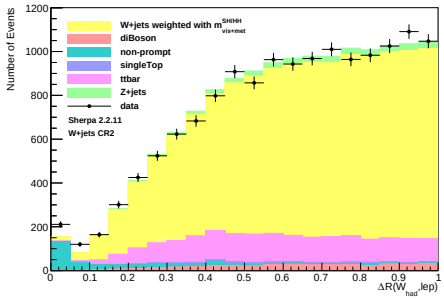
(a) Weighted with lepton p_T (SHERPA 2.2.1)



(b) Weighted with lepton p_T (SHERPA 2.2.11)



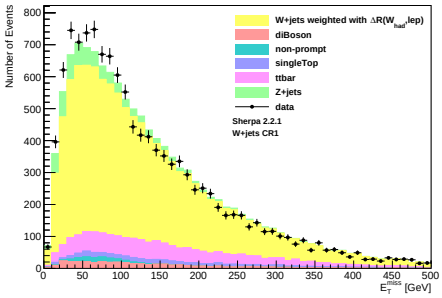
(c) Weighted with $m^{SH/HH}_{\text{vis+met}}$ (SHERPA 2.2.1)



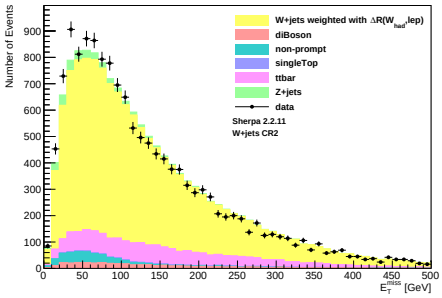
(d) Weighted with $m^{SH/HH}_{\text{vis+met}}$ (SHERPA 2.2.11)

Figure A.6.: The histograms with the stacked background processes for $\Delta R(W_{\text{had}}, \ell)$ in the W+jets CR1 (left plots) and W+jets CR2 (right plots).

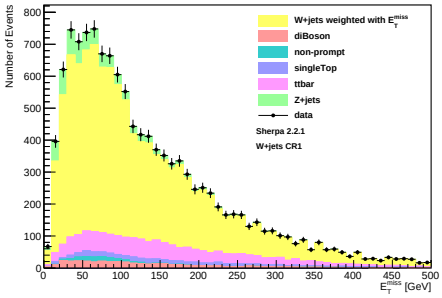
A.2.2. E_T^{miss}



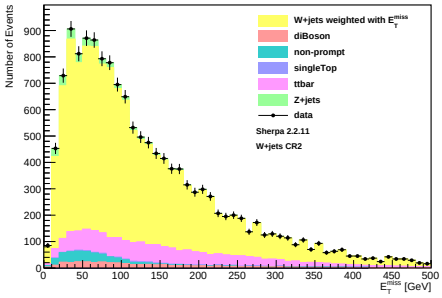
(a) Weighted with $\Delta R(W_{\text{had}}, \ell)$ (Sherpa 2.2.1)



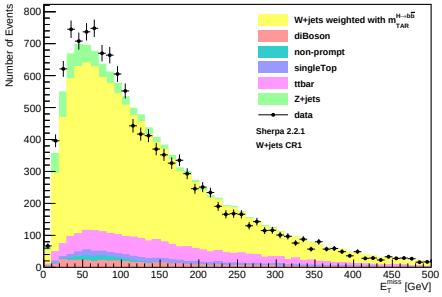
(b) Weighted with $\Delta R(W_{\text{had}}, \ell)$ (Sherpa 2.2.11)



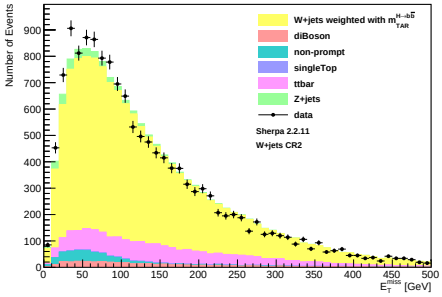
(c) Weighted with E_T^{miss} (Sherpa 2.2.1)



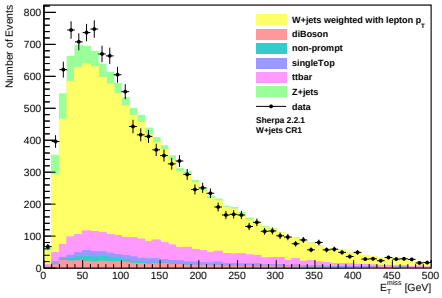
(d) Weighted with E_T^{miss} (Sherpa 2.2.11)



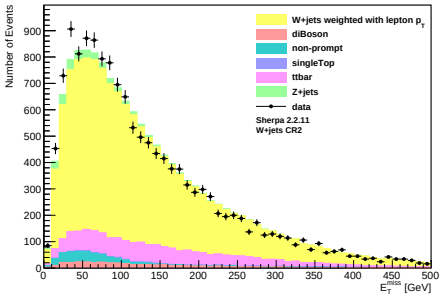
(e) Weighted with $m_{\text{TAR}}^{H\to b\bar{b}}$ (Sherpa 2.2.1)



(f) Weighted with $m_{\text{TAR}}^{H\to b\bar{b}}$ (Sherpa 2.2.11)

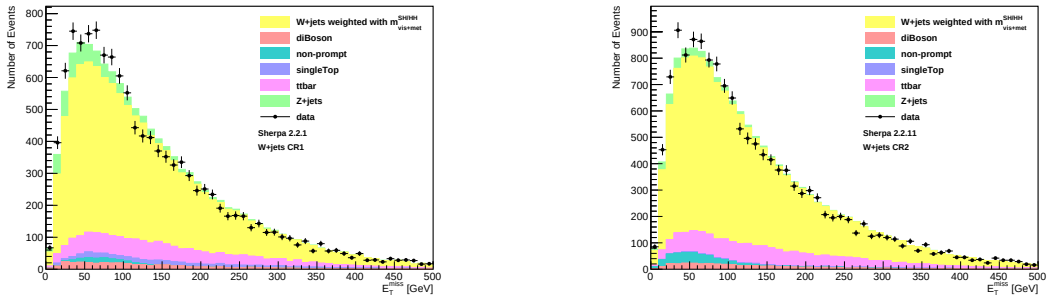


(g) Weighted with lepton p_T (Sherpa 2.2.1)



(h) Weighted with lepton p_T (Sherpa 2.2.11)

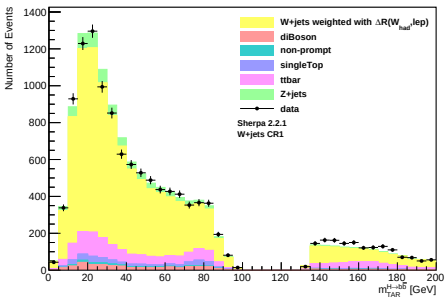
Figure A.7.: The histograms with the stacked background processes for E_T^{miss} in the W+jets CR1 (left plots) and W+jets CR2 (right plots).



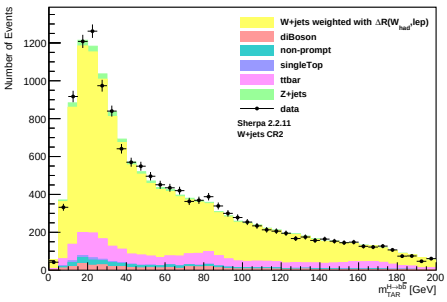
(a) Weighted with $m_{\text{vis+met}}^{SH/HH}$ (SHERPA 2.2.1) ***(b)*** Weighted with $m_{\text{vis+met}}^{SH/HH}$ (SHERPA 2.2.11)

Figure A.8.: The histograms with the stacked background processes for E_T^{miss} in the W+jets CR1 (left plots) and W+jets CR2 (right plots).

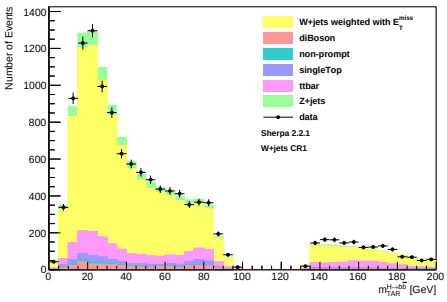
A.2.3. $m_{\mathrm{TAR}}^{H\to b\bar{b}}$



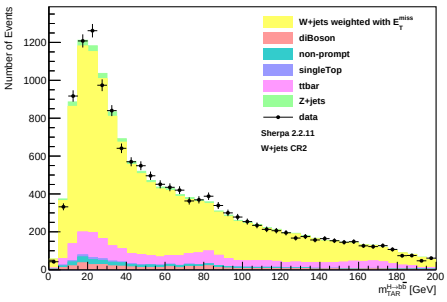
(a) Weighted with $\Delta R(W_{\mathrm{had}}, \ell)$ (SHERPA 2.2.1)



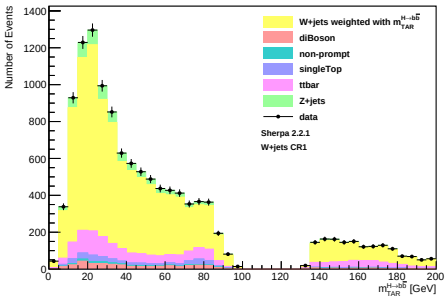
(b) Weighted with $\Delta R(W_{\mathrm{had}}, \ell)$ (SHERPA 2.2.11)



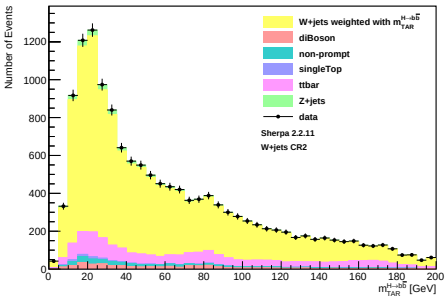
(c) Weighted with E_T^{miss} (SHERPA 2.2.1)



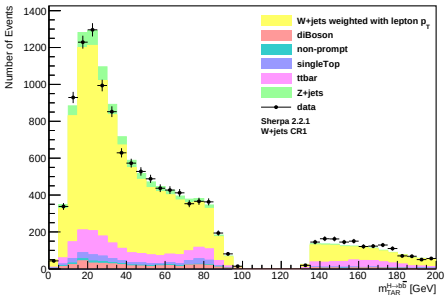
(d) Weighted with E_T^{miss} (SHERPA 2.2.11)



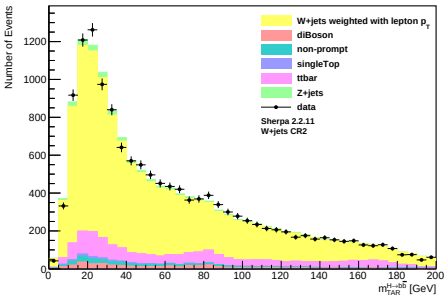
(e) Weighted with $m_{\mathrm{TAR}}^{H\to b\bar{b}}$ (SHERPA 2.2.1)



(f) Weighted with $m_{\mathrm{TAR}}^{H\to b\bar{b}}$ (SHERPA 2.2.11)

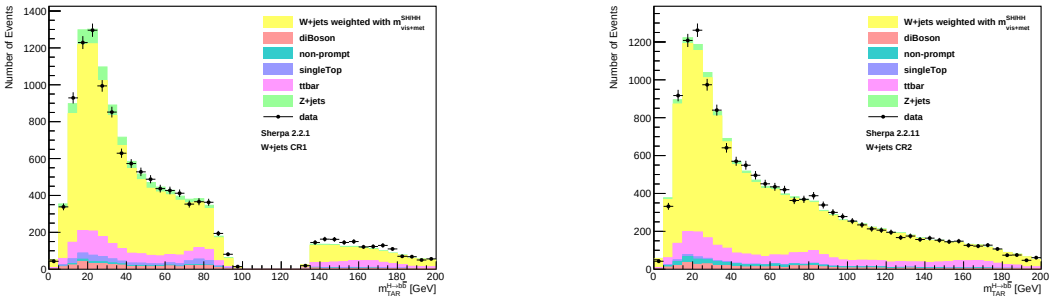


(g) Weighted with lepton p_T (SHERPA 2.2.1)



(h) Weighted with lepton p_T (SHERPA 2.2.11)

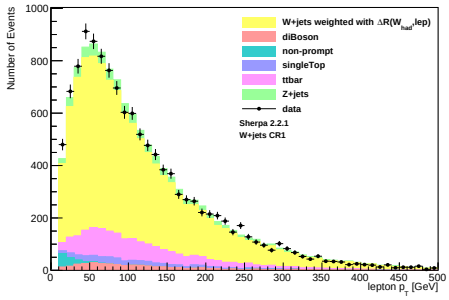
Figure A.9.: The histograms with the stacked background processes for $m_{\mathrm{TAR}}^{H\to b\bar{b}}$ in the W+jets CR1 (left plots) and W+jets CR2 (right plots).



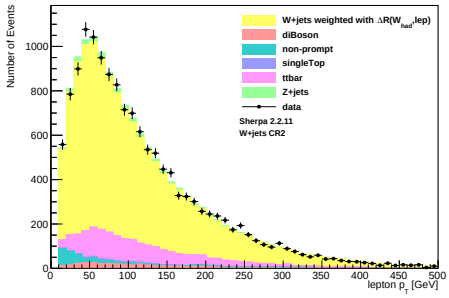
(a) Weighted with $m_{\text{vis+met}}^{SH/HH}$ (SHERPA 2.2.1) ***(b)*** Weighted with $m_{\text{vis+met}}^{SH/HH}$ (SHERPA 2.2.11)

Figure A.10.: The histograms with the stacked background processes for $m_{\text{TAR}}^{H\to b\bar{b}}$ in the W+jets CR1 (left plots) and W+jets CR2 (right plots).

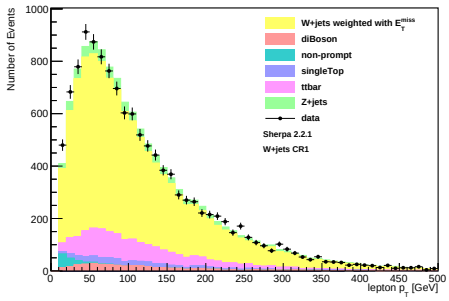
A.2.4. lepton p_T



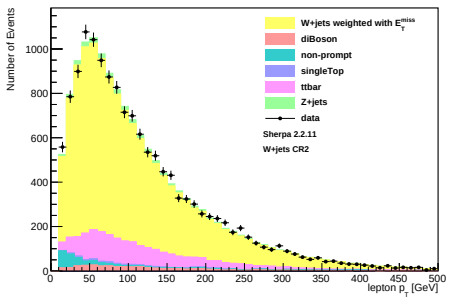
(a) Weighted with $\Delta R(W_{\text{had}}, \ell)$ (SHERPA 2.2.1)



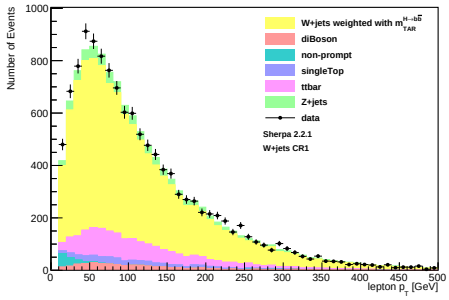
(b) Weighted with $\Delta R(W_{\text{had}}, \ell)$ (SHERPA 2.2.11)



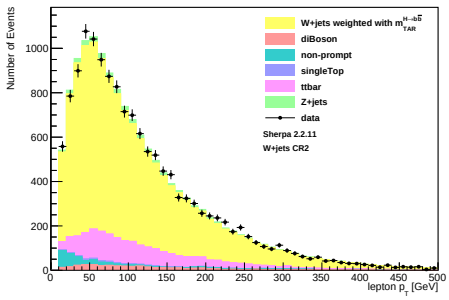
(c) Weighted with E_T^{miss} (SHERPA 2.2.1)



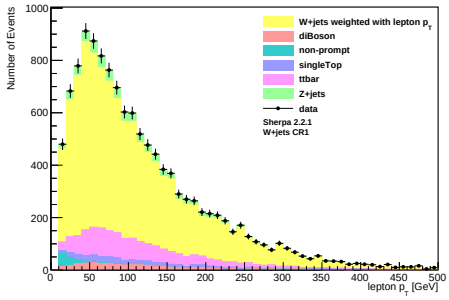
(d) Weighted with E_T^{miss} (SHERPA 2.2.11)



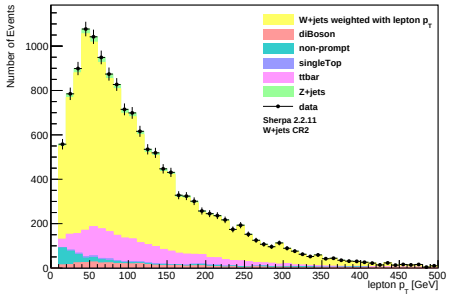
(e) Weighted with $m_{\text{TAR}}^{H\to b\bar{b}}$ (SHERPA 2.2.1)



(f) Weighted with $m_{\text{TAR}}^{H\to b\bar{b}}$ (SHERPA 2.2.11)

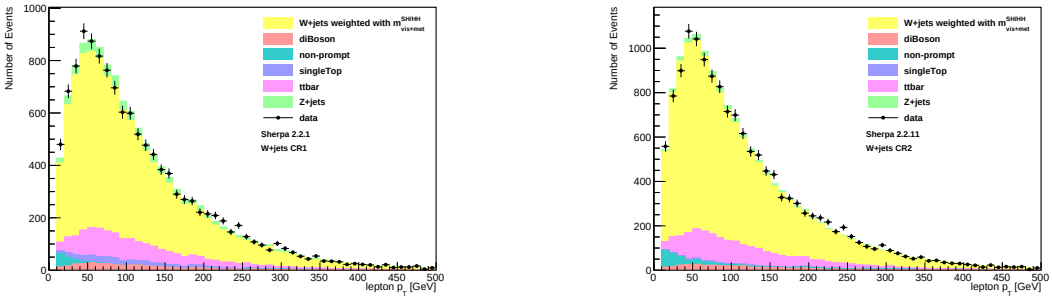


(g) Weighted with lepton p_T (SHERPA 2.2.1)



(h) Weighted with lepton p_T (SHERPA 2.2.11)

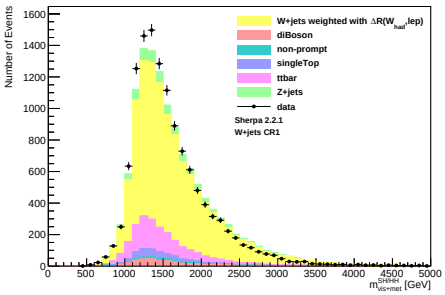
Figure A.11.: The histograms with the stacked background processes for lepton p_T in the W+jets CR1 (left plots) and W+jets CR2 (right plots).



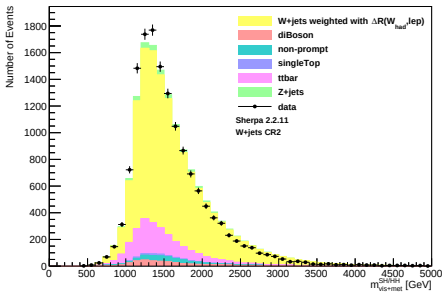
(a) Weighted with $m_{\text{vis+met}}^{SH/HH}$ (SHERPA 2.2.1) ***(b)*** Weighted with $m_{\text{vis+met}}^{SH/HH}$ (SHERPA 2.2.11)

Figure A.12.: The histograms with the stacked background processes for lepton p_T in the W+jets CR1 (left plots) and W+jets CR2 (right plots).

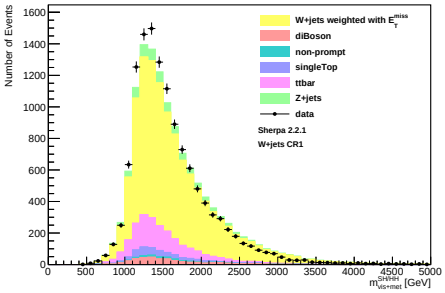
A.2.5. $m_{\text{vis+met}}^{SH/HH}$



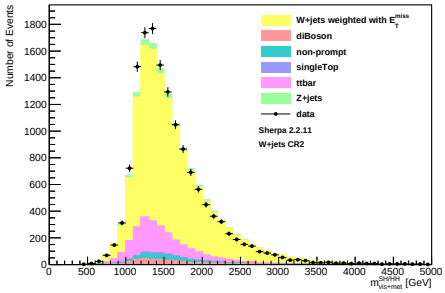
(a) Weighted with $\Delta R(W_{\text{had}}, \ell)$ (SHERPA 2.2.1)



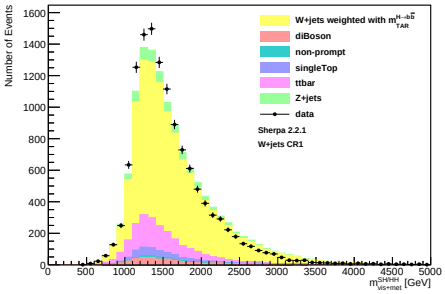
(b) Weighted with $\Delta R(W_{\text{had}}, \ell)$ (SHERPA 2.2.11)



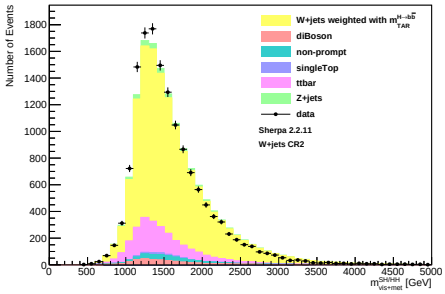
(c) Weighted with E_T^{miss} (SHERPA 2.2.1)



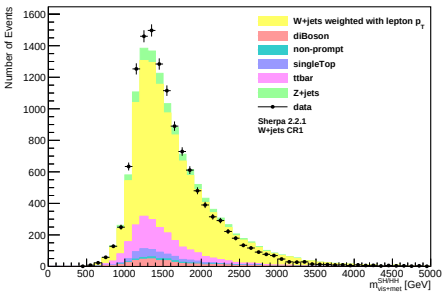
(d) Weighted with E_T^{miss} (SHERPA 2.2.11)



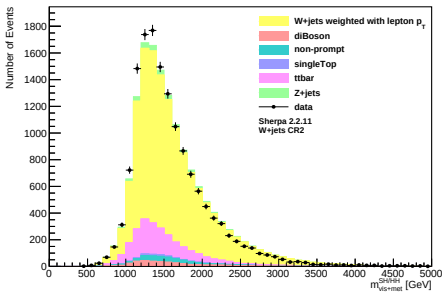
(e) Weighted with $m_{\text{TAR}}^{H\to b\bar{b}}$ (SHERPA 2.2.1)



(f) Weighted with $m_{\text{TAR}}^{H\to b\bar{b}}$ (SHERPA 2.2.11)

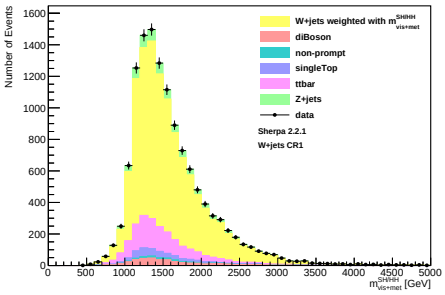


(g) Weighted with lepton p_T (SHERPA 2.2.1)

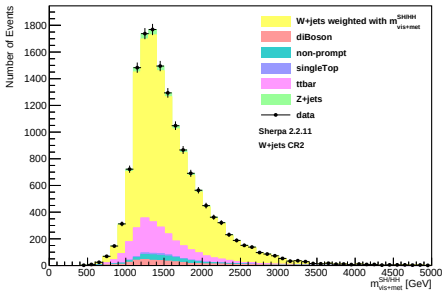


(h) Weighted with lepton p_T (SHERPA 2.2.11)

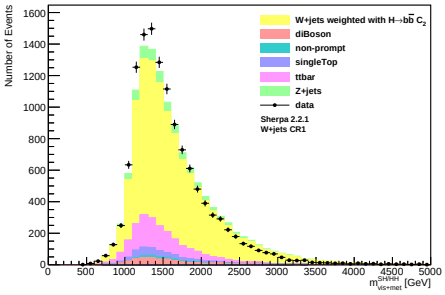
Figure A.13.: The histograms with the stacked background processes for $m_{\text{vis+met}}^{SH/HH}$ in the W+jets CR1 (left plots) and W+jets CR2 (right plots).



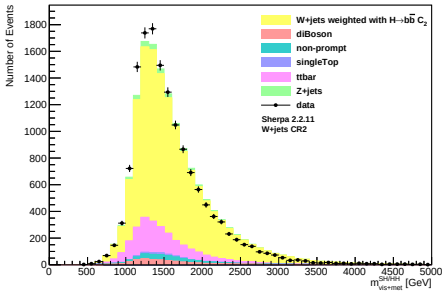
(a) Weighted with $m^{SH/HH}_{\text{vis+met}}$ (SHERPA 2.2.1)



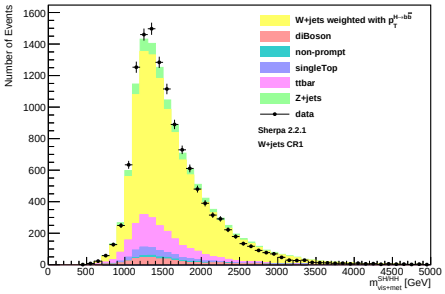
(b) Weighted with $m^{SH/HH}_{\text{vis+met}}$ (SHERPA 2.2.11)



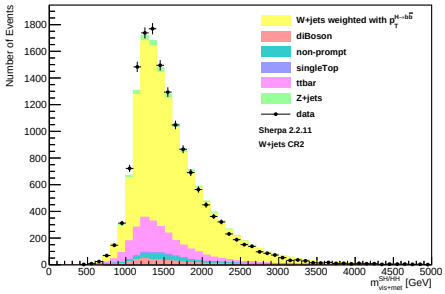
(c) Weighted with $H \to b\bar{b}$ C_2 (SHERPA 2.2.1)



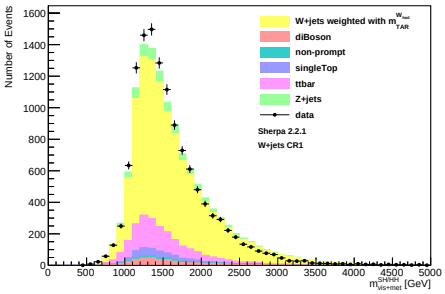
(d) Weighted with $H \to b\bar{b}$ C_2 (SHERPA 2.2.11)



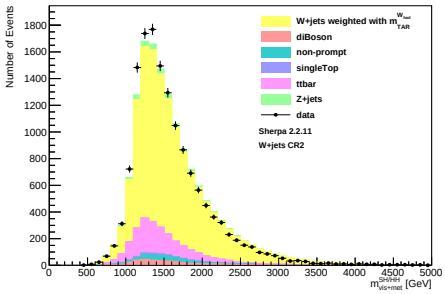
(e) Weighted with $p_T^{H\to b\bar{b}}$ (SHERPA 2.2.1)



(f) Weighted with $p_T^{H\to b\bar{b}}$ (SHERPA 2.2.11)

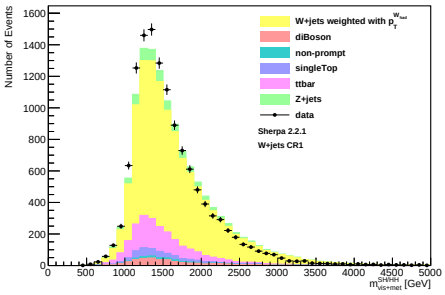


(g) Weighted with $m^{\text{W}_{\text{had}}}_{\text{TAR}}$ (SHERPA 2.2.1)

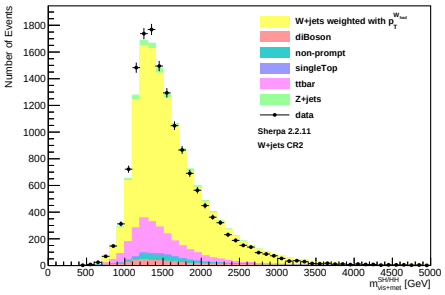


(h) Weighted with $m^{\text{W}_{\text{had}}}_{\text{TAR}}$ (SHERPA 2.2.11)

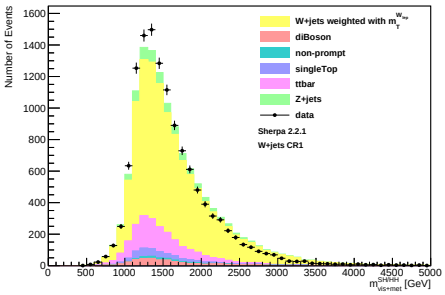
Figure A.14.: The histograms with the stacked background processes for $m^{SH/HH}_{\text{vis+met}}$ in the W+jets CR1 (left plots) and W+jets CR2 (right plots).



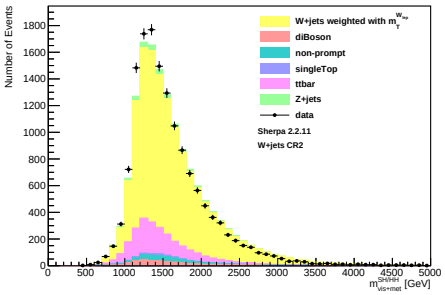
(a) Weighted with $p_T^{W_{\mathrm{had}}}$ (SHERPA 2.2.1)



(b) Weighted with $p_T^{W_{\mathrm{had}}}$ (SHERPA 2.2.11)



(c) Weighted with $m_T^{W_{\mathrm{lep}}}$ (SHERPA 2.2.1)



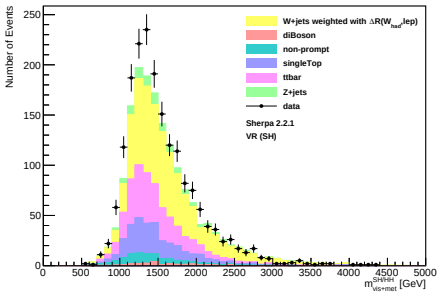
(d) Weighted with $m_T^{W_{\mathrm{lep}}}$ (SHERPA 2.2.11)

Figure A.15.: The histograms with the stacked background processes for $m_{\mathrm{vis+met}}^{SH/HH}$ in the W+jets CR1 (left plots) and W+jets CR2 (right plots).

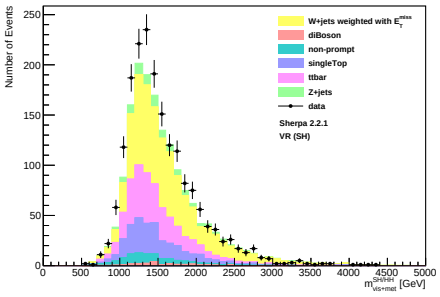
B. $m_{\text{vis+met}}^{SH/HH}$ in the VRs

Here, "Weighted with [variable]" means that the W+jets events are weighted with $\frac{\text{Data}-\text{MC}_{\text{else}}}{\text{MC}_{W+\text{jets}}}$ obtained from the distribution of the variable in the respective W+jets CR.

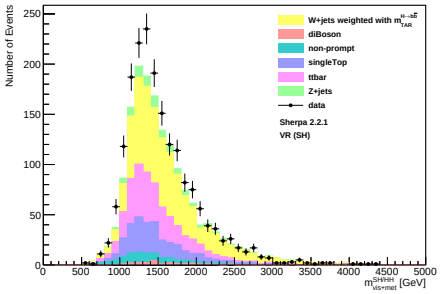
B.1. Weighted W+jets Events in the VR



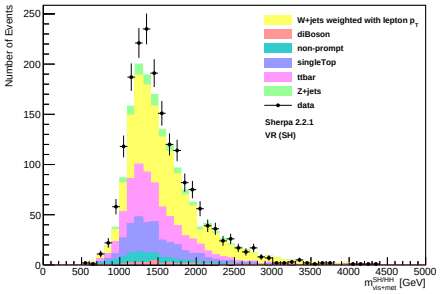
(a) Weighted with $\Delta R(W_{\text{had}}, \ell)$



(b) Weighted with E_T^{miss}

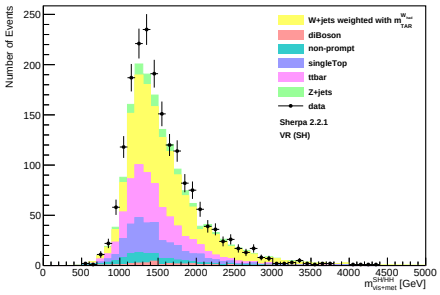


(c) Weighted with $m_{\text{TAR}}^{H\to b\bar{b}}$

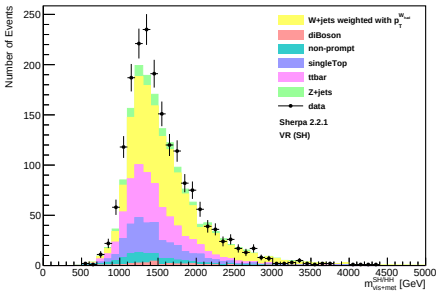


(d) Weighted with lepton p_T

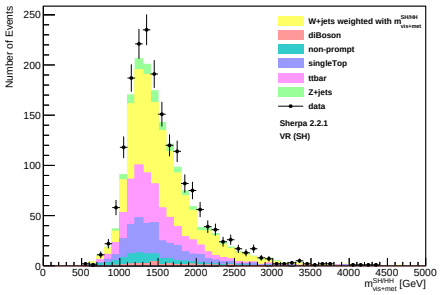
Figure B.1.: The SHERPA 2.2.1 background modelling of $m_{\text{vis+met}}^{SH/HH}$ in the VR (SH selection) where the W+jets background is weighted.



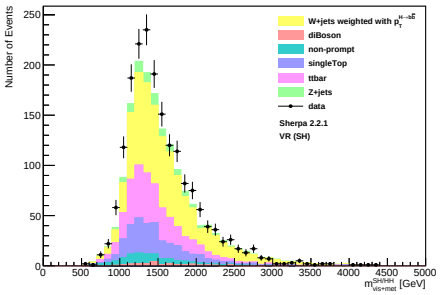
(a) Weighted with $m_{\text{TAR}}^{\text{W}_{\text{had}}}$



(b) Weighted with $p_T^{W_{\text{had}}}$

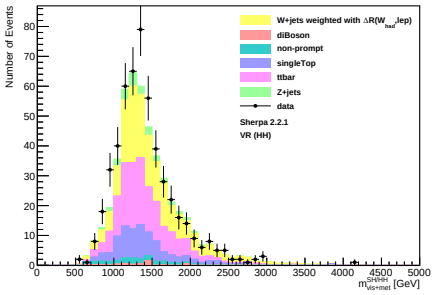


(c) Weighted with $m_{\text{vis+met}}^{SH/HH}$

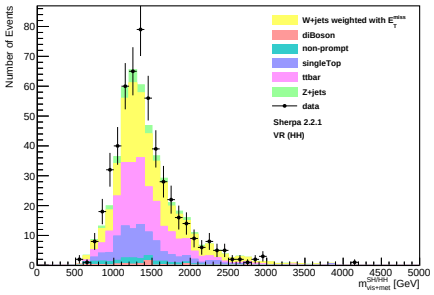


(d) Weighted with $p_T^{H\to b\bar{b}}$

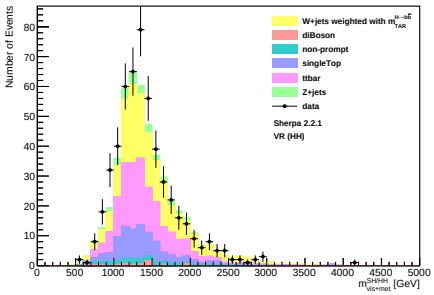
Figure B.2.: The SHERPA 2.2.1 background modelling of $m_{\text{vis+met}}^{SH/HH}$ in the VR (SH selection) where the W+jets background is weighted.



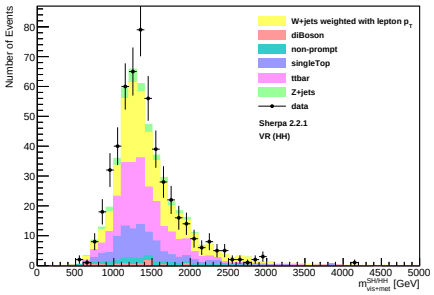
(a) Weighted with $\Delta R(W_{\text{had}}, \ell)$



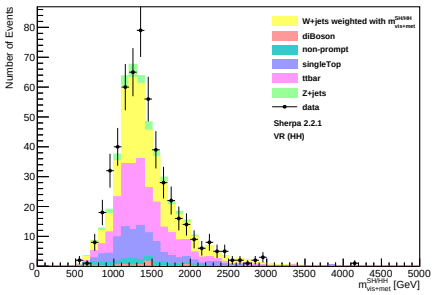
(b) Weighted with E_T^{miss}



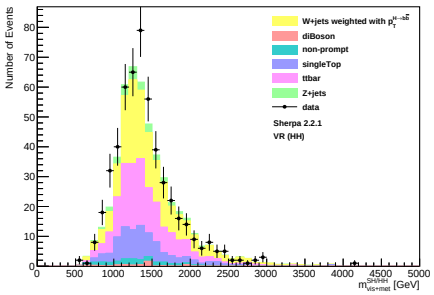
(c) Weighted with $m_{\text{TAR}}^{H\to b\bar{b}}$



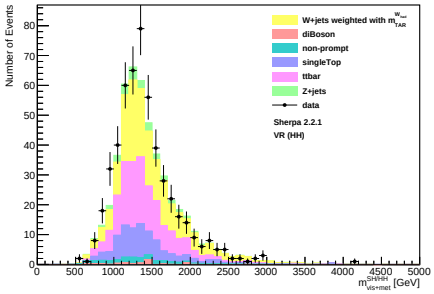
(d) Weighted with lepton p_T



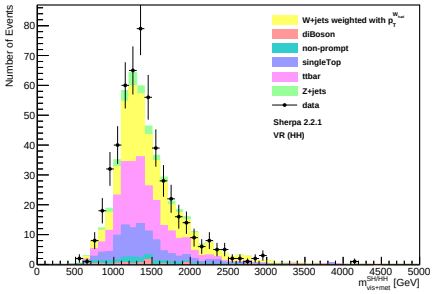
(e) Weighted with $m_{\text{vis+met}}^{SH/HH}$



(f) Weighted with $p_T^{H\to b\bar{b}}$



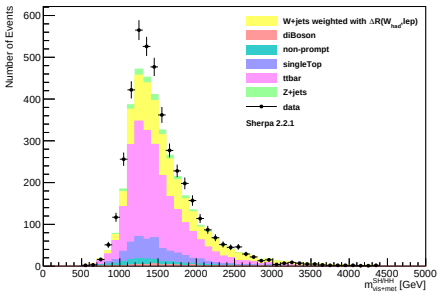
(g) Weighted with $m_{\text{TAR}}^{\text{W}_{\text{had}}}$



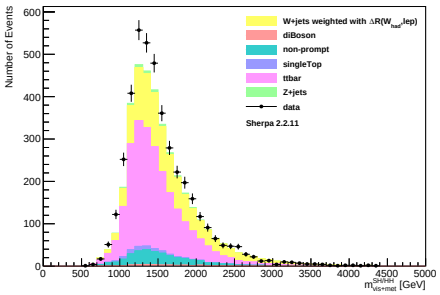
(h) Weighted with $p_T^{W_{\text{had}}}$

Figure B.3.: The SHERPA 2.2.1 background modelling of $m_{\text{vis+met}}^{SH/HH}$ in the VR (HH selection) where the W+jets background is weighted.

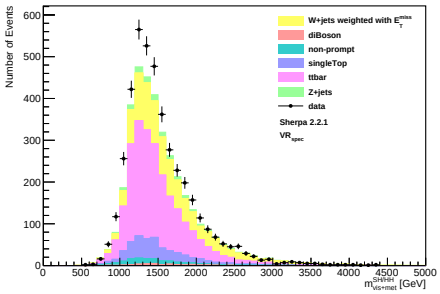
B.2. $m_{\text{vis+met}}^{SH/HH}$ in the VR_{spec}



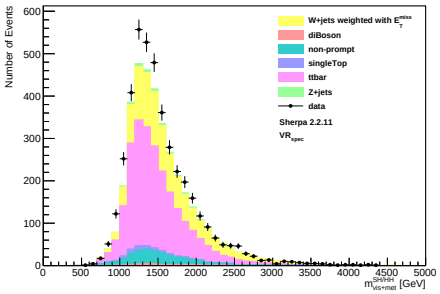
(a) Weighted with $\Delta R(W_{\text{had}}, \ell)$ (SHERPA 2.2.1)



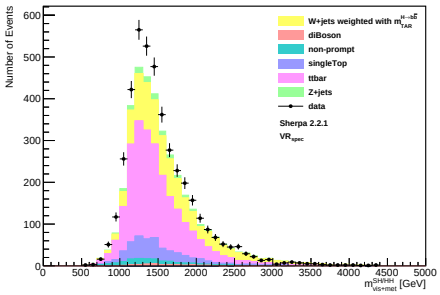
(b) Weighted with $\Delta R(W_{\text{had}}, \ell)$ (SHERPA 2.2.11)



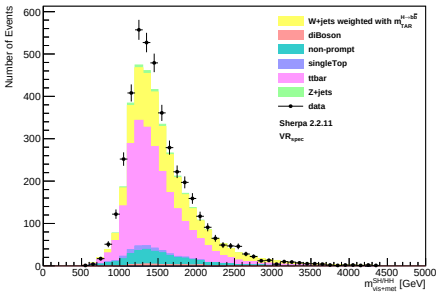
(c) Weighted with E_T^{miss} (SHERPA 2.2.1)



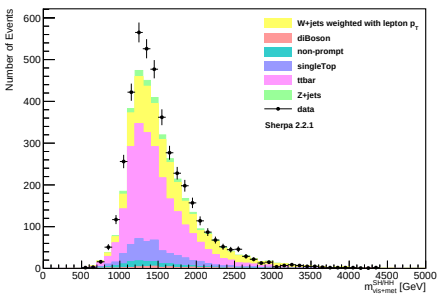
(d) Weighted with E_T^{miss} (SHERPA 2.2.11)



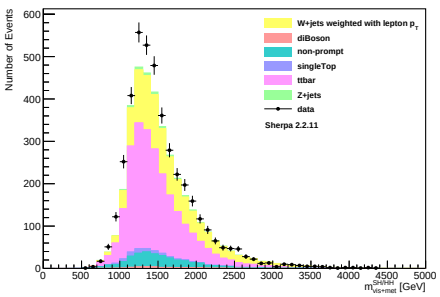
(e) Weighted with $m_{\text{TAR}}^{H\to b\bar{b}}$ (SHERPA 2.2.1)



(f) Weighted with $m_{\text{TAR}}^{H\to b\bar{b}}$ (SHERPA 2.2.11)

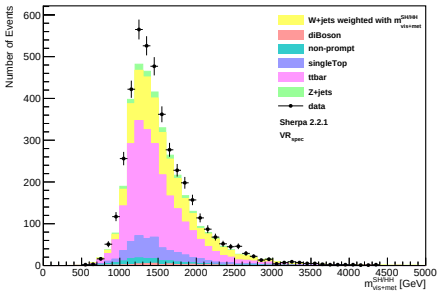


(g) Weighted with lepton p_T (SHERPA 2.2.1)

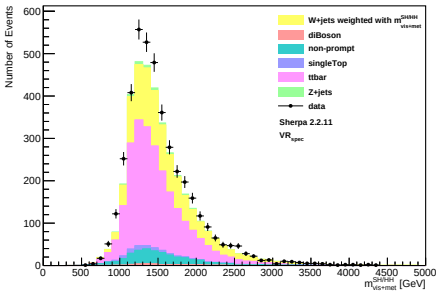


(h) Weighted with lepton p_T (SHERPA 2.2.11)

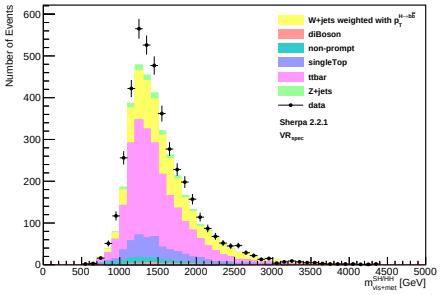
Figure B.4.: The SHERPA 2.2.1 and SHERPA 2.2.11 background modelling of $m_{\text{vis+met}}^{SH/HH}$ in the VR_{spec} where the W+jets background is weighted.



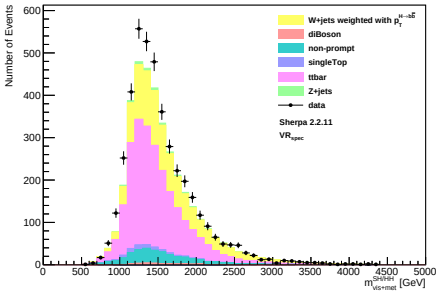
(a) Weighted with $m_{\text{vis+met}}^{SH/HH}$ (SHERPA 2.2.1)



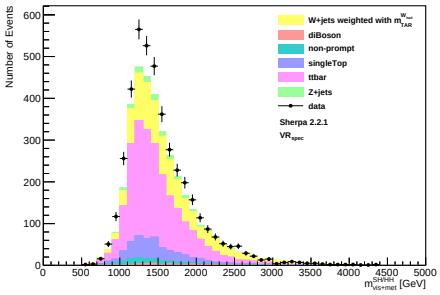
(b) Weighted with $m_{\text{vis+met}}^{SH/HH}$ (SHERPA 2.2.11)



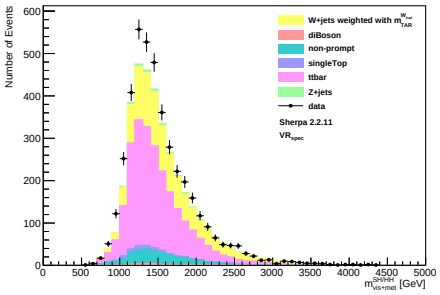
(c) Weighted with $p_T^{H\to b\bar{b}}$ (SHERPA 2.2.1)



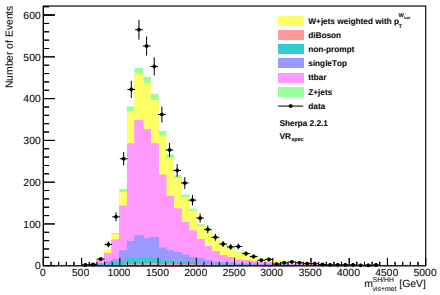
(d) Weighted with $p_T^{H\to b\bar{b}}$ (SHERPA 2.2.11)



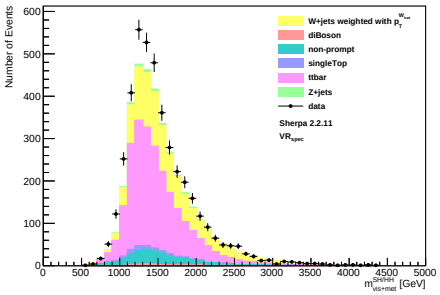
(e) Weighted with $m_{\text{TAR}}^{\text{W}_{\text{had}}}$ (SHERPA 2.2.1)



(f) Weighted with $m_{\text{TAR}}^{\text{W}_{\text{had}}}$ (SHERPA 2.2.11)



(g) Weighted with $p_T^{W_{\text{had}}}$ (SHERPA 2.2.1)



(h) Weighted with $p_T^{W_{\text{had}}}$ (SHERPA 2.2.11)

Figure B.5.: The SHERPA 2.2.1 and SHERPA 2.2.11 background modelling of $m_{\text{vis+met}}^{SH/HH}$ in the VR_{spec} where the W+jets background is weighted.

Bibliography

[1] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19**, 1264 (1967)

[2] S. L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22**, 579 (1961)

[3] A. Salam, *Weak and Electromagnetic Interactions*, ed. Nobel Symposium No. 8 (Almqvist & Wiksell, Stockholm, 1968)

[4] A. Salam, J. Ward, *Electromagnetic and weak interactions*, Phys Lett. **13(2)**, 168 (1964)

[5] L. Evans, P. Bryant, *LHC Machine*, JINST **3(08)**, S08001 (2008)

[6] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3(08)**, S08003 (2008)

[7] CMS Collaboration, *The CMS experiment at the CERN LHC*, JINST **3(08)**, S08004 (2008)

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

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

[10] B. A. Robson, *The Matter-Antimatter Asymmetry Problem*, JHEPGC **04(01)**, 166 (2018)

[11] R. L. Workman, et al. (Particle Data Group), *Review of Particle Physics*, Progress of Theoretical and Experimental Physics **2022(8)** (2022), 083C01

[12] H. D. Politzer, *Asymptotic Freedom: An Approach to Strong Interactions*, Phys. Rept. **14**, 129 (1974)

[13] W. Hollik, *Quantum field theory and the Standard Model*, in *2009 European School of High-Energy Physics* (2010), 1012.3883

[14] ATLAS Collaboration, *Measurement of the W-boson mass in pp collisions at* $\sqrt{s}$ = 7 TeV *with the ATLAS detector*, EPJ C **78(2)** (2018)

[15] The LEP Collaborations, *Measurement of the mass of the Z boson and the energy calibration of LEP*, Phys. Lett. B **307(1)**, 187 (1993)

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

[17] CMS Collaboration, *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, Phys. Lett. B **716**, 30 (2012)

[18] M. Thomson, *Modern Particle Physics*, Cambridge University Press (2013)

[19] B. Povh, et al., *Teilchen und Kerne: Eine Einführung in die physikalischen Konzepte (Springer-Lehrbuch) (German Edition)*, Springer Spektrum (2014)

[20] D. de Florian, et al. (LHC Higgs Cross Section Working Group), *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, CERN Yellow Reports: Monographs, CERN, Geneva (2017), URL `https://cds.cern.ch/record/2227475`

[21] B. Di Micco, et al., *Higgs boson pair production at colliders: status and perspectives. Higgs boson potential at colliders: Status and perspectives*, Rev. Phys. **5**, 100045. 161 p (2020)

[22] G. C. Branco, P. M. Ferreira, L. Lavoura, et al., *Theory and phenomenology of two-Higgs-doublet models*, Phys. Rept. **516**, 1 (2012)

[23] T. Robens, T. Stefaniak, J. Wittbrodt, *Two-real-scalar-singlet extension of the SM: LHC phenomenology and benchmark scenarios*, The European Physical Journal C **80(2)**, 151 (2020)

[24] B. Hespel, D. Lopez-Val, E. Vryonidou, *Higgs pair production via gluon fusion in the Two-Higgs-Doublet Model*, JHEP **09**, 124 (2014)

[25] E. Mobs, *The CERN accelerator complex in 2019. Complexe des accélérateurs du CERN en 2019* (2019), general Photo, URL `https://cds.cern.ch/record/2684277`

[26] R. Alemany-Fernandez, et al., *Operation and Configuration of the LHC in Run 1* (2013), URL `https://cds.cern.ch/record/1631030`

[27] J. Wenninger, *Operation and Configuration of the LHC in Run 2* (2019), URL `https://cds.cern.ch/record/2668326`

[28] ATLAS Collaboration, *Production and integration of the ATLAS Insertable B-Layer*, JINST **13(05)**, T05008 (2018)

[29] ATLAS Collaboration, *Particle Identification Performance of the ATLAS Transition Radiation Tracker*, Technical Report ATLAS-CONF-2011-128, CERN, Geneva (2011), URL `http://cds.cern.ch/record/1383793`

[30] ATLAS Collaboration, *Performance of the ATLAS Trigger System in 2015*, Eur. Phys. J. C **77(5)**, 317 (2017)

[31] K. Abeling, *Search for resonant Higgs boson pair production in the $b\bar{b}WW^*$ decay channel in the boosted 1-lepton final state using the full Run 2 ATLAS dataset*, Ph.D. thesis, II.Physik-UniGö-Diss-2022/01

[32] ATLAS Collaboration, *Search for resonant boosted HH and SH production in the $bbVV$ decay channel with 0 or 1 lepton in the final state using the full Run 2 ATLAS data*, Technical Report ATL-COM-PHYS-2022-020, CERN, Geneva (2022), URL `https://cds.cern.ch/record/2799424`

[33] J. Hammersley, *Monte Carlo Methods*, Springer (2011)

[34] S. Weinzierl, *Introduction to Monte Carlo methods* (2000), `hep-ph/0006269`

[35] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}$ =13 TeV*, Eur. Phys. J. C **76(5)**, 292 (2016)

[36] ATLAS Collaboration, *Working points*, URL `http://twiki.cern.ch/twiki/bin/view/AtlasProtected/MCPAnalysisGuidelinesMC16#Working_points`. [Last accessed: 2023-01-26]

[37] ATLAS Collaboration, *Muon working points*, URL `https://twiki.cern.ch/twiki/bin/view/AtlasProtected/RecommendedIsolationWPs#Muon_working_points`. [Last accessed: 2023-01-26]

[38] M. Cacciari, G. P. Salam, G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04**, 063 (2008)

[39] O. Brandt, et al., *Track assisted techniques for jet substructure*, Technical Report ATL-COM-PHYS-2018-455, CERN, Geneva (2018), URL `https://cds.cern.ch/record/2315709`

[40] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, Eur. Phys. J. C **77(7)**, 466 (2017)

[41] ATLAS Collaboration, *Performance of b-Jet Identification in the ATLAS Experiment*, JINST **11(04)**, P04008 (2016)

[42] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **79(11)**, 970 (2019)

[43] ATLAS Collaboration, *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run*, Technical Report ATL-PHYS-PUB-2017-013, CERN, Geneva (2017), URL `https://cds.cern.ch/record/2273281`

[44] Sherpa Collaboration, *Event Generation with Sherpa 2.2*, SciPost Phys. **7(3)**, 034 (2019)

[45] J. R. Taylor, *An Introduction to Error Analysis*, University Science Books (1997)

Acknowledgement

First of all, I would like to thank you Prof. Stan Lai for giving me the opportunity to write my bachelor thesis in your research group. Without you, this would have not been possible and I only have endless gratitude left for you. Thank you for your support and for believing in me.

I would also like to thank Kira. You were the one that helped me with every problem I was not able to solve on my own and answered every question precisely. Thank you for being so patient and the one, I could always come to with anything. I will miss our little conversations about god and the world after our private research chats.

Also, I would like to thank Prof. Schumann for being my second referee.

Last but not least, thank you to my family and my friends. Thank you for your endless support.

Erklärung nach §13(9) der Prüfungsordnung für den Bachelor-Studiengang Physik und den Master-Studiengang Physik an der Universität Göttingen:
Hiermit erkläre ich, dass ich diese Abschlussarbeit selbständig verfasst habe, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe und alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen wurden, als solche kenntlich gemacht habe.
Darüberhinaus erkläre ich, dass diese Abschlussarbeit nicht, auch nicht auszugsweise, im Rahmen einer nichtbestandenen Prüfung an dieser oder einer anderen Hochschule eingereicht wurde.

Göttingen, den 28. März 2023

(Zobeyer Ghafoor)