Department of Physics and Astronomy
University of Heidelberg

Bachelor Thesis in Physics
submitted by

Katharina Julia Bews

born in Esslingen am Neckar (Germany)

2021

Estimating $W \to \tau\nu_\tau$ background processes in dark matter searches with energetic jets and missing energy

This bachelor thesis has been carried out by Katharina Julia Bews at the
Kirchhoff Institute for Physics in Heidelberg
under the supervision of
Prof. Dr. Monica Dunford

Abstract

In this bachelor thesis the $W \to \tau\nu_\tau$ background process is studied with the ATLAS detector in order to correct the normalization of the Monte Carlo predicted $W \to \tau\nu_\tau$ background process in the signal region, which is assigned to the signature of the dark matter production. This correction is performed using a scale factor derived with the single τ-lepton control region. In addition, the transverse W-boson mass as well as the τ-lepton transverse momentum are constrained and the stability of the scale factor will be shown. Furthermore, systematic uncertainties based on the reconstruction of the τ-lepton are applied and their influence on the scale factor as well as their significance compared to the statistical uncertainties are analyzed.

Kurzfassung

In dieser Bachelorarbeit wird der $W \to \tau\nu_\tau$ Hintergrundprozess mit dem ATLAS-Detektor untersucht, um die Normalisierung des Monte Carlo vorhergesagten $W \to \tau\nu_\tau$ Hintergrundprozesses in der Signalregion zu korrigieren. Diese Korrektur wird unter Verwendung eines Skalierungsfaktors durchgeführt, der aus der entsprechenden Kontrollregion abgeleitet wird. Dabei wird zusätzlich das Spektrum der transversalen W-Boson Masse, sowie des transversalen τ-Lepton Impulses auf unterschiedliche Weise eingeschränkt und die entsprechend veränderten Ergebnisse des Skalierungsfaktors miteinander verglichen. Weiterhin werden die zur Rekonstruktion des τ-Leptons gehörigen systematischen Unsicherheiten studiert und deren Einfluss auf den Skalierungsfaktor, als auch deren Größenordnung verglichen mit den statistischen Unsicherheiten, analysiert.

Contents

1 Introduction

1.1 The Standard Model of Particle Physics

The Standard Model of particle physics (SM) describes our current knowledge of all elementary particles and the non-gravitational interactions between them. The following information were mainly taken from [1].

1.1.1 Fundamental Particles

Fundamental Particles can be divided into two groups: fermions and bosons, which can be differentiated by their spin. The fundamental particles of the SM are summarized in figure 1.1.

Particles with half-integer spins are called fermions. In the SM there are twelve types of fermions, divided into six quarks and six leptons. They are classified according to whether they interact via the strong force or not. In general, the SM is separated in three generations with two particles of each classification.

Leptons are grouped in pairs of a negatively charged and an electric neutral particle per generation where the later ones are known as neutrinos. They do not interact strongly, but weakly. The charged fermions additionally interact electromagnetically.

Quarks are separated in up-type and down-type quark per generation distinguishable by their different charges. Up-type quarks have a charge of $\frac{2}{3}$ and down-type quarks an electric charge of $-\frac{1}{3}$. Unlike the leptons, quarks do interact through their color charge strongly, electromagnetically and weakly.

To each fermion belongs a corresponding anti-particle, which has the same mass and spin, but has an opposite electromagnetic color and charge.

Bosons can also be split into two different groups. The gauge bosons consist of four particles with integer spin 1, while the Higgs boson is the only scalar boson with an integer spin of 0. Whereas gauge bosons are force carriers, the Higgs boson describes why the other particles, except for the gluon and the photon, are massive through the interaction with the so-called Higgs mechanism.

1.1.2 Fundamental Forces

The SM includes three of the fundamental forces in nature, excluding the gravitational force, where each of them can be explained by a field theory.

The theory of quantum electrodynamics describes the interactions of charged particles by the exchange of a virtual gauge boson, namely the photon (γ). The photon is uncharged and massless and couples only to charged particles, which is necessary

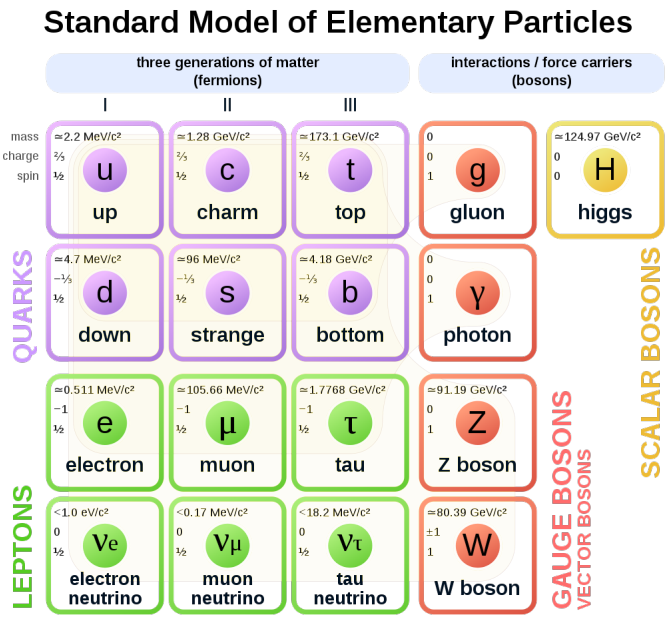


Figure 1.1: *Overview of the fundamental particles in the Standard Model of particle physics, taken from [2].*

for the electromagnetic force. The strength of the force decreases with $\frac{1}{r}$, whereby r defines the distance, and has an infinite range.

The strong force is described by the theory of quantum chromodynamics and is the strongest force in the SM. Quarks are the only particles that are able to interact via the strong force because they carry color charge. Due to their color charge, quarks are never observed as free particles but are confined to bound color neutral states called hadrons. Hadrons can be divided into two categories: mesons and baryons. Mesons consist of two quarks (an quark and anti-quark) and baryons consists of three quarks (three quarks of different color which leads to neutral color charge). The associated gauge bosons are the gluons. Eight massless and electrically neutral gluons exist that are able to couple to color charge. The increasing strength of the strong force with the distance makes the hadron production possible.

The weak force is explained by the electroweak theory. The gauge bosons of the weak force are the $W^{\pm}$- and Z^0-bosons, which couple via hypercharge resulting in the possible interaction with all fermions. The $W^{\pm}$-boson is the only gauge boson that is allowed to produce a change of quark flavor. Since these bosons are very massive, the weak force is suppressed.

1.2 The ATLAS Experiment

The Large Hadron Collider (LHC) is with an extension of 27 kilometers the world's largest and highest-energy particle collider, built by the European Organisation of Nuclear Research (CERN) in Geneva. Bunches of protons will be induced to collide in one of the four particle detectors (ATLAS, CMS, LHCb and ALICE) to provide proton-proton collision which results can be studied afterwards. This thesis

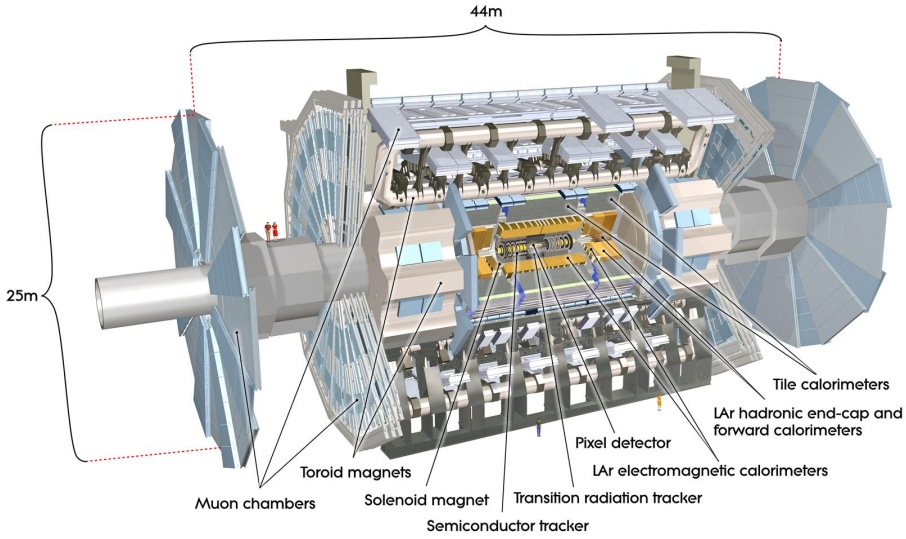


Figure 1.2: *Overall layout of the ATLAS detector [3]*

will cover the data of proton-proton collisions recorded by the ATLAS detector in 2015/16 with a center of mass energy of $\sqrt{s} = 13$ TeV and a total integrated luminosity of 36.2 fb^{-1}. The information due to the ATLAS detector are taken from [4].

The ATLAS detector is a cylindrical particle detector with a length of 44 meters, a height of 25 meters and a weight of 7000 tonnes. The coordinate system, describing the detector, defines the interaction point as the origin of the coordinate system. Additionally, it sets the z-axis in the direction of the beam line and the x-y plane transverse to it. The positive x-axis points from the origin of the coordinate system to the center of the LHC ring and the y-axis points upwards. Two angles ϕ and θ will be defined to completely describe the detector geometry. The azimuthal angle ϕ describes the x-y plane and the polar angle θ corresponds to the angle of the beam axis.

An ATLAS detector layout is given in figure 1.2. The main structure of the detector includes the inner detector, the calorimeter system and the muon chambers. Each of these systems are shrouded by a barrel and a pair of end-caps.

The inner detector records precise information about the position of traversing particles, accomplished through several concentric layers, including pixel and silicon microstrip trackers used in combination with the Transition Radiation Tracker (TRT). A charged particle can be detected by its deposition of energy through its ionization with the detector material. The whole system is embedded into a magnetic field generated by a solenoid magnet. This magnetic field deflects charged particles, which makes it possible to derive the charge and momentum of the particle.

The inner detector and the solenoid magnet are located at the center of the detector. Further outwards is the calorimeter system, separated into the electromagnetic and hadronic calorimeter. They are designed to measure the energy and direction of a particle. Unlike the inner system, it is possible to measure charged and neutral particles in the calorimeter system.

The electromagnetic calorimeter consists of a lead absorption layer. The two end-caps, along with the barrel, consist of accordeon-shaped kapton electrodes filled with liquid argon used as active material. With the electromagnetic calorimeter and the inner detector it is possible to separate an electron from a photon, due to the charge of the electron. Both particles will be stopped in the calorimeter and the deposit energy can be measured. The interaction of electrons with the detector material leads to bremsstrahlung, whereas photons are involved in pair production. In the pair production photons form electron pairs, which subsequently form photons in turn by bremsstrahlung. This is repeated several times and electromagnetic showers occur.

The hadronic calorimeter consists of a steel absorption layer and scintillating tiles. Both of the end-caps are filled with liquid argon. The hadronic calorimeter enables the detection and identification of hadrons. It is thicker than the electromagnetic calorimeter to compensate the smaller retardation of the hadrons. Hadrons interact with the absorber material and cascades of secondary low-energy particles emerge, which ionize the active material and cause hadronic showers. Hadronic showers can be explained by quarks which are torn apart by the collision until there is enough interaction energy to create new quark pairs. This so-called process of hadronization repeats itself and a large number of hadrons is created. The direction of flight of the resulting particles is determined by the direction of the initial particle, therefore all particles fly within a beam in the same direction, the so-called jet.

The muon chambers with their toroid magnets form the last layer of the detector. The active medium is a gas mixture of carbon-dioxide and argon. The constituents of the chambers are the Monitored Drift Tubes and Cathode Strip Chambers for tracking purposes and the Resistive Plate Chambers and Thin Gap Chambers for triggering purposes. Muons can only be minimally ionized, which is why they are deflected by the toroid magnets in order to identify their momentum by the curve they leave in the chamber due to their charge.

Since neutrinos only interact weakly, they cannot be detected. Energy conservation gives information about missing transverse energy, which hints to invisible particles like neutrinos.

1.3 Physics Beyond the Standard Model: Dark Matter

The Standard Model is a widely approved and successful theory. Nonetheless, it still does not have an answer to all questions, as for example the abundance of dark matter in the universe. Different astrophysical observations already revealed evidences for dark matter, but yet only through the gravitational force [5]. Dark matter does not interact with the electromagnetic field and therefore not with photons, which makes it difficult to detect. There is no possible candidate in the SM, except the neutrino, which would fit to that description.

One dark matter candidate is the weakly interacting massive particle (WIMP). Like neutrinos, WIMPs are invisible particles and will not be detected by the LHC detector. Therefore, dark matter processes can only be indirectly detected by mostly

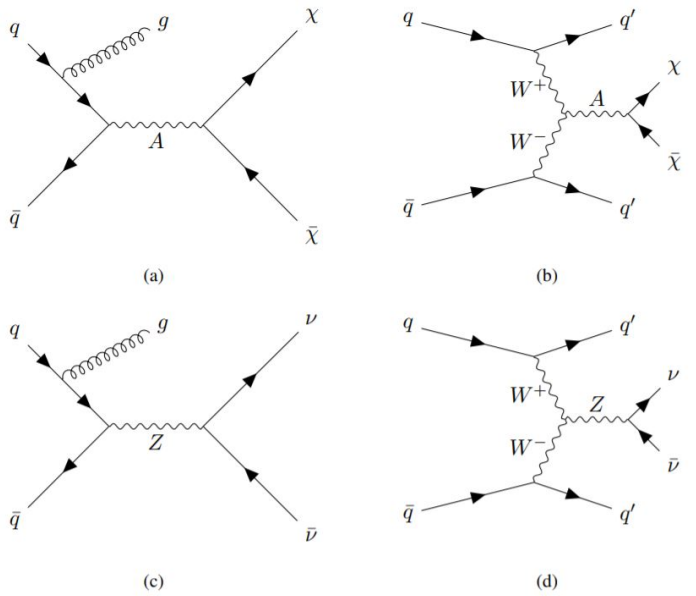


Figure 1.3: *Feynman diagrams for the production of WIMPs with one jet in (a) and vector-boson fusion in (b). Similar to that, the Feynman diagrams in the SM for $Z \rightarrow \nu\overline{\nu}$ with one jet in (c) and vector-boson fusion in (d) [6].*

hadronic particle jets as a result of their production process and missing transverse momentum p_T^{miss} that should exist due to momentum conservation. p_T^{miss} is a measured momentum imbalance in the plane perpendicular to the beam direction and it can be used to allocate invisible particles.

Two processes exist explaining the production of WIMPs. In the first process a mediator decays into a WIMP pair ($\chi\overline{\chi}$) and a gluon-initiated jet is produced, as shown in 1.3a. In principle, this decay can occur without the mentioned jet, but it would be difficult to identify this process at all, which is why it is explicitly searched for a jet. Since there is at least one detectable high energy jet, this process will be called mono-jet decay process. In the second process the production occurs via vector-boson fusion (VBF), as shown in 1.3b. In this mechanism two quarks scatter off each other at low angle leading to virtual vector bosons, that can either be W- or Z-bosons. These vector bosons scatter and produce the mediator A, which decays into WIMPs, while simultaneously two jets are produced [6].

Another dominant Standard Model process leads to the same final states, as visible in 1.3c and 1.3d, and is carried out by the Z-boson decaying into a pair of neutrinos ($\nu\overline{\nu}$). Consequently, it is difficult to test the theory of the dark matter candidates since both processes will not be detected in the detector and need to be distinguished.

2 Analysis

2.1 Analysis Strategy

2.1.1 The Search for Dark Matter

A promising signature for the search for dark matter is missing transverse energy E_T^{miss}, an energy imbalance in the transverse plane of the detector. E_T^{miss} does not differ from p_T^{miss} because the mass, which otherwise plays a role in calculating the energy, is negligible at high energies. Two signatures that can result from different dark matter channels are chosen to be analyzed, the mono-jet and VBF signature, following from the decays visible in figures 1.3a and 1.3b. The dark matter production signature, consisting of a mono-jet or dijet and E_T^{miss}, is assigned to the signal region.

As it is visible in figure 1.3c and 1.3d, the signature of $Z \to \nu\overline{\nu}$ does not differ from the signature of the dark matter production, since the neutrinos, like the WIMPs, interact weakly and are therefore invisible. This leads to a signal region including both processes. As the background of the $Z \to \nu\overline{\nu}$ process is irreducible, it is necessary to measure an exact cross section of this decay process and search for deviations from the Standard Model.

Nevertheless, there are other background processes that can have an effect on the signal region. Any decay process in which particles are misidentified or not detected at all might lead to incorrectly classified processes. Leptons, which are not within the range of the detector acceptance, are misidentified as a hadronic jet or are not reconstructed cannot be rejected by the signal region. Especially $W \to l\nu_l$ decays, where l denotes a lepton, but also leptonically decaying Z-bosons having an identical signature as $Z \to \nu\overline{\nu}$ except of the charge and the mass of the fermions, contribute to the background of the signal region. However, the background contribution of these Z-boson decays is significant lower in the signal region than the leptonically W-boson decay, because two leptons instead of one need to be lost or misidentified. Additionally, single top, diboson, triboson and other processes, where several bosons are produced (EW), contribute to the background of the signal region.

The background expectation is simulated via Monte Carlo simulation and are compared to the observation. Nonetheless, the simulations still have significant uncertainties and do not describe the processes correctly. Control regions for the most relevant background process are implemented and created to be as similar as possible to the signature in the signal region, so that a strong correlation between the systematic uncertainties of the the two regions exist. There are control regions, which constraints the systematic uncertainties and propagates them in the signal region, whereas in other control regions the normalization of the Monte Carlo prediction is determined to correct the normalization in the signal region.

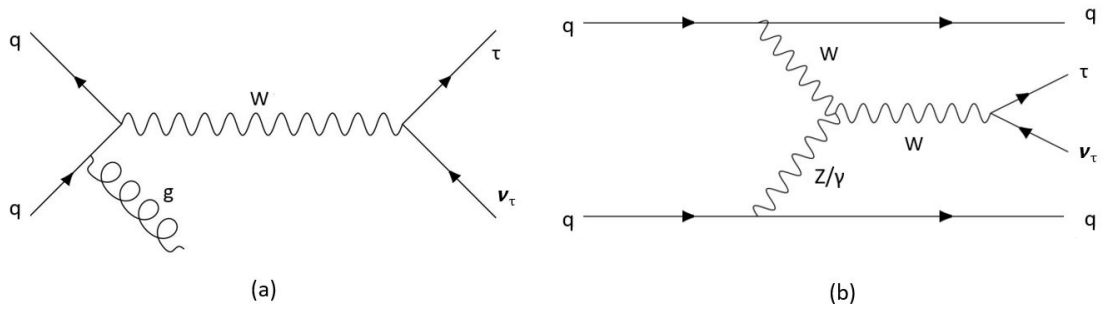


Figure 2.1: *Feynman diagrams of the $W \to \tau\nu_\tau$ decay process. The mono-jet signature is visible in (a) and the VBF signature in (b). The Feynman diagrams were created with [7].*

2.1.2 $W \to \tau\nu_\tau$ Analysis

In this thesis the $W \to \tau\nu_\tau$ background process will be studied in order to correct the Monte Carlo normalization of its background contribution in the signal region. Both signatures, mono-jet and vector-boson fusion will be discussed. A possible Feynman diagram for each signature of the $W \to \tau\nu_\tau$ decay is illustrated in figure 2.1. The Feynman diagrams are, except for the additional lepton, similar to the processes in figure 1.3. Therefore, a misidentification or non-detection of the τ-lepton would wrongly sort the process into the signal region.

Among the W-boson backgrounds in the signal region, the $W \to \tau\nu_\tau$ process has the largest contribution, which can be illustrated with the following example. In figure 2.2 the measured E_T^{miss} mono-jet distribution in the signal region is displayed along with the sum of all Monte Carlo simulated background processes. A stack histogram is presented with the number of events scaled by their bin width as a function of the missing transverse energy in GeV. The histogram lists up every process contributing to the signal region. All these processes, except the $Z \to \nu\overline{\nu}$ decay, were not identified or reconstructed correctly and were sorted in the signal region. The amount of $W \to \tau\nu_\tau$ is with a percentage around 20% the dominant reducible background. The dominance of this decay can be explained by the short lifetime of the τ-lepton, which complicates a corresponding lepton veto in the signal region. For the other leptons, namely the electron and the muon, this lepton veto exists.

The estimation of the simulation in the signal region does not necessarily correspond to the observation. Background processes can therefore not be accurately assessed. Precisely because the control region can compare the Monte Carlo prediction to the data and thus determine the proportion of agreement, the single τ-lepton control region has an important role. All events with transverse missing energy and a mono-jet or dijet signature as well as a measured τ-lepton are selected. Dominant background processes are the top, the diboson and the triboson decay processes. The most dominant background contribution is the EW production (fig.3.2).

The first step will be to normalize the Monte Carlo prediction by the number of data events in the control region, both corresponding to $W \to \tau\nu_\tau$ and to correct the

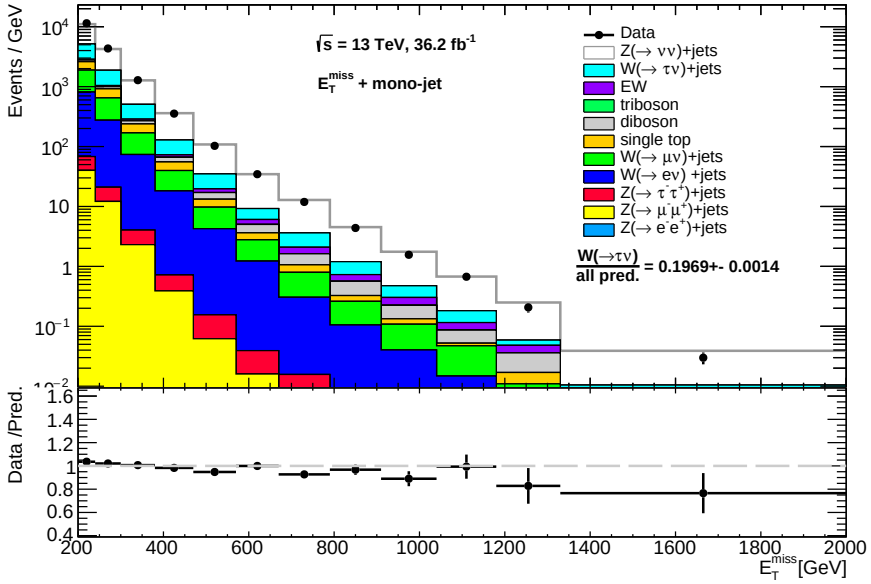


Figure 2.2: *Agreement of the data to Monte Carlo simulation for the mono-jet E_T^{miss} distribution in the signal region. The amount of $W \to \tau\nu_\tau$ is given. Standard modified selection criteria are applied.*

signal region by the calculated normalization, referred to as scale factor. Furthermore, individual modified selection criteria will be applied on the scale factor studies and it will be analyzed how stable the scale factor behaves. These selections include restriction on the permitted range of the transverse momentum of the τ-lepton and of the transverse mass of the W-boson, which will be discussed in chapter 2.5. Another step is to analyze the statistical and systematic uncertainties of the $W \to \tau\nu_\tau$ distribution. Thereby, systematic uncertainties regarding to the reconstruction of the τ-lepton will be compared.

2.2 Monte Carlo Simulation

Monte Carlo simulation provides simulated events from the Standard Model and cross sections of decay processes. This is useful for estimating background contributions as well as for the calibration of reconstructed objects. Furthermore, it can be used to probe the kinematic of physic models beyond the Standard Model. The information in the following is taken from [8] and [9].

The Monte Carlo event simulation proceeds in two stages: The event generation and the detector simulation.

The event generation begins with a hard-scattering proton-proton collision and the calculation of the corresponding matrix element, which describes the probability and the strength of the transition between an initial and a final state. With the parton distribution function the components and the momentum fraction of the proton at a specific energy scale can be determined. This stage is called parton-level. Finally, in the particle-level partons hadronize by an implemented parton shower algorithm and form stable bound states.

In the detector simulation the passage of the final states particles through the ATLAS detector generated in the particle-level and the interaction with it is simulated. This stage is referred to as reconstruction level. The performance of the object reconstruction and the event selection is based on the results of the detector simulation. These results are processed in the same way as real recorded data, inefficiencies of the detector and misidentification are included. Following to that, they can be treated equally. This kind of simulated data were used for the analysis in this thesis.

Different Monte Carlo generators were used to produce simulated samples. SHERPA is a parton shower Monte Carlo generator, which covers all productions with a vector boson and its jet (V +jets), as well as the production of diboson and triboson samples. SHERPA generates samples with up to three final state jets. The parton distribution function used is NNPDF3.0NNLO, which is a next-to-next-to-leading order (NNLO) generator. Additionally, single top samples are generated using PYTHIA and POWHEG with NNPDF3.0NLO, which is a next-to-leading order (NLO) generator.

Leading-order (LO) means that only first order Feynman-diagram processes will be considered. NLO and NNLO consider the second and the third order respectively of Feynman-diagrams by including loops.

2.3 Physics Objects Reconstruction

Following to the event generation, the digitization and reconstruction process take place just before the data can be analyzed. The digitization thereby converts the detector response into a format compatible with the real output, which afterwards can be reconstructed. In the following, the references [10] and [11] were used.

Jets

In a hard interaction quarks and gluons are produced and hadronize immediately. Due to their confinement they cannot be detected as free particles. During the hadronization numerous hadronic particles are produced, which all originate from the same initial quark or gluon and can be clustered into a jet.

The basis of the calorimeter system in the detector consists of a three-dimensional topological cluster, which is composed of calorimeter cells. One of the calorimeter cells is identified with a certain energy significance above noise level and subsequently neighboring cells with a similar significance are detected forming a circular cluster. The noise is thereby composed by electronic and pile-up noise. In the next step the topological cluster is split and individual clusters are calibrated by local cluster weighting calibration (LCW). This allows a classification of the jets and enables to build them using the anti-kt algorithm.

Taus

The τ-lepton is the heaviest charged lepton in the Standard Model and decays into other charged leptons and hadrons. With a higher probability it decays hadronically

into one charged pion referred to as one prong decay or into three charged pions (three prong decay) or kaons. Pions are the simplest mesons in the Standard Model, consisting of an up and down quark, whereas kaons consist of an up and strange quark. The one prong and three prong decays are accompanied by neutral pions and neutrinos. In ATLAS, the τ reconstruction is concentrated on the hadronically decay of τ-leptons. The identification considers jets with their anti-kt topological clusters and searches for deviations in clusters and tracks around the jet direction, which are sent trough a τ identification algorithm. Identified τ-lepton decays will be calibrated using LCW.

Missing Transverse Energy

The reconstruction process of E_T^{miss} uses the energy deposit in the calorimeter cells and sums up particle tracks in order to include low-momentum particles. Calorimeter cells, which are associated with reconstructed high momentum objects and cells without assigned objects are taken into account. The E_T^{miss} is then determined by the negative sum of the momenta of the muon tracks and the negative sum of the calibrated cell energy of these contributions projected onto the x and y direction.

2.4 Event Selection

Events will be assigned to a signal region or multiple control regions regarding their reconstructed final states.

The signal region is characterized by a transverse missing energy and jet or dijet signature, additionally a lepton veto is applied. In the $W \to l\nu$ control regions an additional lepton is required, whereas in the $Z \to l\bar{l}$ control regions two opposite charged leptons are needed. Since processes beyond the Standard Model produce more events with transverse missing energy and have a higher energy range, due to a high energy scale and a large mediator mass, a range of $E_T^{miss} > 200$ GeV has to be adopted, as well as an invariant dijet mass of $m_{jj} > 200$ GeV. The cross section of the mono-jet and VBF process is measured in bins of E_T^{miss}. Additionally, the VBF process is determined in bins of invariant dijet mass m_{jj} and the azimuthal angle between the two leading jets $\Delta\phi_{jj}$.

A single high energy jet and a large amount of transverse missing energy form the signature of the mono-jet process. The mono-jet process has a higher cross section than the VBF process and therefore a large statistical dataset. The leading jet forms the main contribution to the size of E_T^{miss}, because it largely has to balance the outgoing transverse Z-boson momentum, that is why $p_T > 120$ GeV is expected [5]. Due to the covering of trackers in the detector, the angle relative to the beam axis is set to $|\eta| < 2.4$. An angle between the ϕ direction of the missing energy and the first four leading jets is set larger than 0.4 to suppress the multijet backgrounds.

The VBF process distinguishes by its dijet and transverse missing energy signature. Both jets need to balance the transverse Z-boson momentum, leading to a smaller momentum than for the mono-jet signature. An 80 GeV and 50 GeV cut are applied, since the first p_T jet suppresses the QCD background [5].

The restriction of the lepton momentum for $W \rightarrow l\nu$ processes correspond to $p_T > 30$ GeV. The leading lepton momentum for $Z \rightarrow l\bar{l}$ processes is selected to be $p_T > 80$ GeV to suppress the non leptonic backgrounds. The second leading lepton momentum is set to be lower than 7 GeV. The invariant mass m_l for the two leptons lies at 7 GeV.

The Z-boson invariant mass is constrained for values between 66 GeV and 116 GeV. The standard selection for the transverse mass M_T of the W-boson lies between 30 GeV and 100 GeV.

In addition, the transverse W-boson mass and the transverse τ-lepton momentum shall be studied in the single τ-lepton control region for different ranges.

2.5 Modified Selection Criteria

For the results in section 3.2, different modified selection criteria were adjusted on the W-boson transverse mass M_T and the τ-lepton transverse momentum p_T.

The τ-lepton transverse momentum is used because protons do not necessarily have the same momentum before the proton-proton collision since they could already have been decelerated or scattered. Therefore, the momentum along the beam axis is not zero, but the transverse momentum is because only particles which were created during the collision are taken into account.

The W-boson decays leptonically including a non-detectable neutrino, which makes it difficult to determine the W-boson mass directly. Therefore, the transverse mass will be introduced as the following

$$M_T^2(W) = 2p_T(l)p_T^{miss}(\nu)[1 - cos(\Delta\phi)]. \tag{2.1}$$

The transverse mass depends on the transverse momentum of the τ-lepton $p_T(l)$, the missing transverse momentum of the neutrino $p_T^{miss}(\nu)$ and the angle between the lepton and the neutrino in the transverse plane $\Delta\phi$ [12].

The parameters M_T and p_T were chosen to be constrained because they influence the signal region regarding the τ-lepton background following from the $W \rightarrow \tau\nu_\tau$ decay. The control region should correspond to the signal region as closely as possible. Since the τ-lepton signature is the only difference between both regions, the

Selection names	W-boson transverse mass M_T	τ transverse momentum p_T
Standard Selection	30-100 GeV	> 30 GeV
Restricted M_T Selection	40-90 GeV	
Unrestricted M_T Selection	full spectrum: 0-2000 GeV	
Restricted p_T Selection	30-100 GeV	> 40 GeV
Restricted p_T & M_T Selection	40-90 GeV	

Table 2.1: Definition of the used modified selection criteria depending on the W-boson transverse mass and the τ-lepton transverse momentum.

normalization of the Monte Carlo predicted $W \rightarrow \tau\nu_\tau$ process can be determined in the control region and applied to the corresponding normalization in the signal region (section 3.2). A varying M_T or p_T range has an effect on the normalization of the Monte Carlo predictions. Therefore, it is important to know how stable the normalization in the control region is, it should not depend on the different selections.

An overview of the applied selections is given in table 2.1. It is well known that the W-boson mass is approximately 80 GeV, but the missing energy from the neutrino makes it difficult to measure the W-mass correctly. The mass spectrum of M_T would not display a mass peak around 80 GeV, but rather a smearing between 30-110 GeV as visible in figure 2.3. To collect the most ideal and useful data, the standard selection is chosen between the area of 30-100 GeV.

The fact that the τ-lepton momentum p_T is influencing the transverse mass of the W-boson, as visible in equation (2.1), is another reason for looking at different selections regarding this parameter. The standard selection for p_T is set to larger than 30 GeV in order to suppress non-leptonically background.

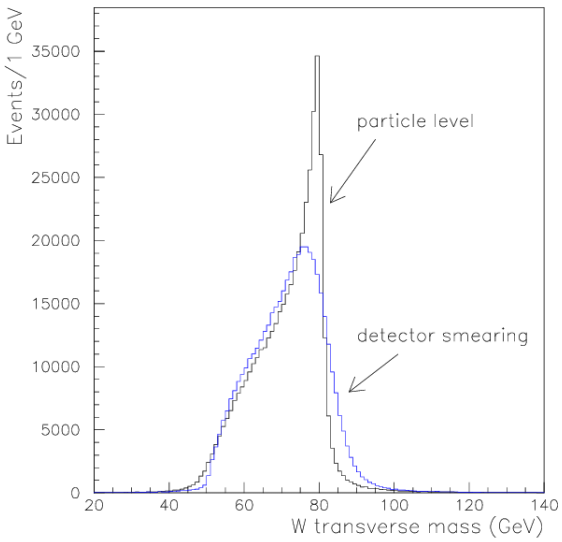


Figure 2.3: *The spectrum of the W-boson transverse mass at particle level (black) and with included energy and momentum resolution leading to detector smearing (blue) [12].*

3 Results

3.1 Analysis of the single τ-lepton Control Region

3.1.1 Comparison of Signal Region to Control Region

The data analyzed here were recorded 2015/16 by the ATLAS detector with a total mass energy of $\sqrt{s} = 13$ TeV and a total integrated luminosity of 36.2 fb^{-1}. This data shall be compared to the reconstruction-level predictions of the Monte Carlo simulation with the aim to get the proportion of the $W \to \tau\nu_\tau$ process and the agreement of the data to the Monte Carlo predictions in the signal and single τ-lepton control region. The ratio between the data and the Monte Carlo simulated $W \to \tau\nu_\tau$ events in the control region will be needed in order to determine a correction factor, which is applied on the $W \to \tau\nu_\tau$ background contribution in the signal region.

Below, multiple histograms are displayed for the mono-jet and VBF E_T^{miss} distribution and the VBF m_{jj} and $\Delta\phi_{jj}$ distribution in the signal region (fig.3.1) and in the control region (fig.3.2). The data is displayed next to a number of Monte Carlo predicted background processes from the Standard Model. The Monte Carlo normalization to the data is determined by the ratio $R_{data/MC}$ of the data events to the Monte Carlo simulated events and is visible in the lower panel of the histograms. The ratio is calculated for each bin as follows

$$R_{data/MC} = \frac{N_{data}}{N_{pred.}^{all}}, \tag{3.1}$$

pred. stands for Monte Carlo prediction events, whereas *data* represents the number of data events. The index *all* includes all generated Standard Model processes.

The uncertainties in the figure correspond to the statistical uncertainties of the data and the Monte Carlo predictions. The statistical uncertainty for the data σ_{data} is assumed to follow a Poisson distribution, therefore the square root of the total number of events N are taken

$$\sigma_{data} = \sqrt{N_{data}}. \tag{3.2}$$

The Monte Carlo prediction follows a Gaussian distribution, the uncertainty of each bin is determined for each individual event by its weight. The weight is applied to make the Monte Carlo simulation more consistent with the expected data and to correct events, which already have weights differing from one due to the most Monte Carlo generators. The uncertainty of the Monte Carlo prediction $\sigma_{pred.}$ can then be calculated by the summation of all events i over all bins j

$$\sigma^2_{pred.} = \sum_{n=1}^{j}(\sum_{n=1}^{i} w_i^2), \tag{3.3}$$

where w indicates the weight. The uncertainty of the ratio $\Delta R_{data/MC}$ from equation (3.1) is introduced as

$$\Big[\frac{\Delta R_{data/MC}}{R_{data/MC}}\Big]^2 = \Big[\frac{\Delta N^{all}_{pred.}}{N^{all}_{pred.}}\Big]^2 + \Big[\frac{\Delta N_{data}}{N_{data}}\Big]^2. \tag{3.4}$$

Multiplying the denominator and taking the square root leads to

$$\Delta R_{data/MC} = R_{data/MC} * \sqrt{\Big[\frac{\sigma^{all}_{pred.}}{N^{all}_{pred.}}\Big]^2 + \Big[\frac{\sigma_{data}}{N_{data}}\Big]^2} \tag{3.5}$$

$$= R_{data/MC} * \sqrt{\Big[\frac{\sigma^{all}_{pred.}}{N^{all}_{pred.}}\Big]^2 + \frac{1}{N_{data}}}. \tag{3.6}$$

Since this thesis tries to correct the amount of $W \to \tau\nu_\tau$ background in the signal region, the relative contribution of this process is given in the figures in order to get an impression of the dimension of the $W \to \tau\nu_\tau$ amount. The relative contribution is calculated by the ratio $R_{W\to\tau\nu_\tau}$ with the number of Monte Carlo $W \to \tau\nu_\tau$ events and the total number of predicted Monte Carlo events in the corresponding region as

$$R_{W\to\tau\nu_\tau} = \frac{N^{W}_{pred.}}{N^{all}_{pred.}}. \tag{3.7}$$

N^W represents, depending on the index, the number of data or prediction events for the $W \to \tau\nu_\tau$ process. The uncertainty of the ratio is calculated similar to equation (3.4) as

$$\Delta R_{W\to\tau\nu_\tau} = R_{W\to\tau\nu_\tau} * \sqrt{\Big[\frac{\sigma^{W}_{pred.}}{N^{W}_{pred.}}\Big]^2 + \Big[\frac{\sigma^{all}_{pred.}}{N^{all}_{pred.}}\Big]^2}. \tag{3.8}$$

The size of the $W \to \tau\nu_\tau$ process lies approximately at 20% for every observable in the signal region and is the largest Monte Carlo predicted background contribution. The statistical uncertainties contribute less than 0.3%. For the single τ-lepton control region a percentage of 71-78% with a statistical error less than 1% can be observed, which indicates that a high amount of predicted $W \to \tau\nu_\tau$ events were correctly sorted into the control region and only a small amount of background contributions exist. Hence, the control region can be well taken to correct the relative background contribution of $W \to \tau\nu_\tau$ processes in the signal region. Since a smaller number of events exists in the VBF region, the statistical uncertainty is slightly larger than in the mono-jet region.

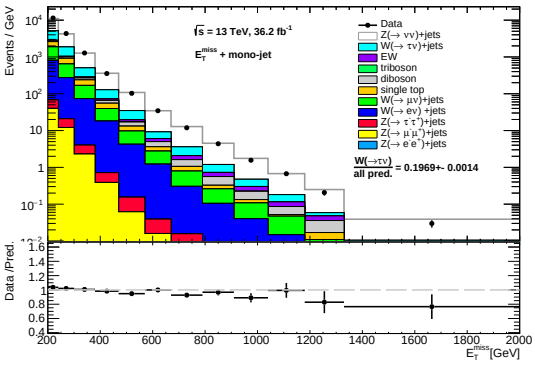
Figure 3.1a and 3.2a show the mono-jet E_T^{miss} distribution in the signal and control

region. Both regions have a Monte Carlo prediction and data agreement of about 5-10% for a low missing transverse energy range. The statistical uncertainties reach only about 1-5%. For high values of E_T^{miss} the ratio decreases in the signal region until a disagreement of 25% is reached, which means that the predictions are overestimated. In addition, the statistical uncertainties become increasingly larger, which is due to proton-proton collisions at higher energies. Since more quarks and gluons are formed by a higher momentum transfer, they suppress the weak interaction and fewer events are counted in total. In the control region the ratio fluctuates until it reaches a disagreement of 80%. The statistical uncertainties rise until a discrepancy of about 20% in the signal region and even 90% in the control region due to a larger number of events in the signal region than in the control region.

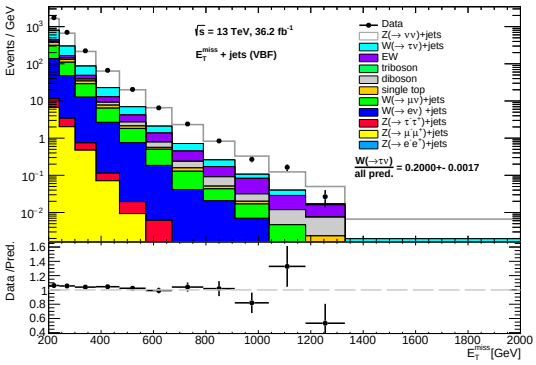
The E_T^{miss} distribution in the VBF signal region (fig.3.1b) shows a general agreement between the data and the Monte Carlo prediction for low values of E_T^{miss} within ranges up to 900 GeV, the distribution is on a scale of 5% relatively flat. A similar course is observed for the control region until a range of 500 GeV (fig.3.2b). Afterwards, the disagreement between the data and the Monte Carlo simulation increases strongly for high E_T^{miss} in both regions. The statistical uncertainties are comparable to the one in the mono-jet signature.

In the m_{jj} VBF signal region (fig.3.1c) the Monte Carlo predictions are overestimated by 10% in the lower m_{jj} range and underestimated by 25% in the upper range. The varying course in the ratio of the data to the prediction can be explained by modeling problems in the Sherpa simulation [13]. In the control region the ratio deviates strongly especially for high m_{jj} ranges. The statistical uncertainties are similar to the VBF E_T^{miss} observable.

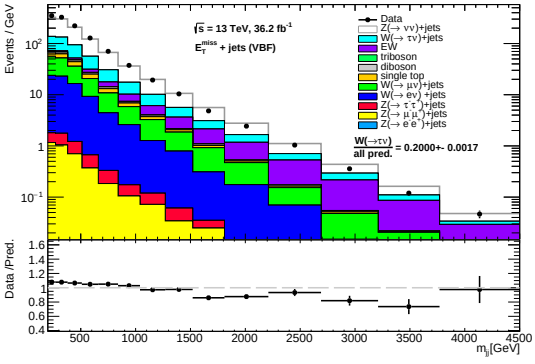
The $\Delta\phi_{jj}$ distribution has a good data to Monte Carlo simulation agreement. The Monte Carlo prediction is underestimated for the signal region up to 15% when $\Delta\phi_{jj}$ tends towards zero. In the control region the data to prediction ratio fluctuates around one and reaches a disagreement up to 20%.



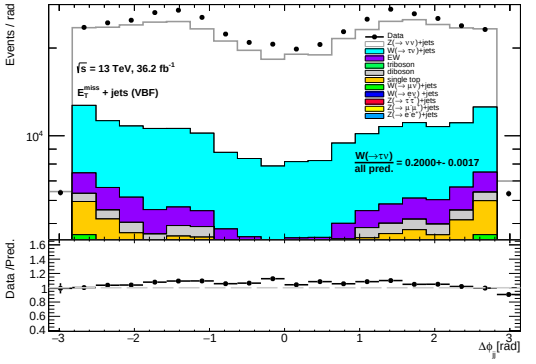
(a) *Mono-jet E_T^{miss} distribution*



(b) *VBF E_T^{miss} distribution*

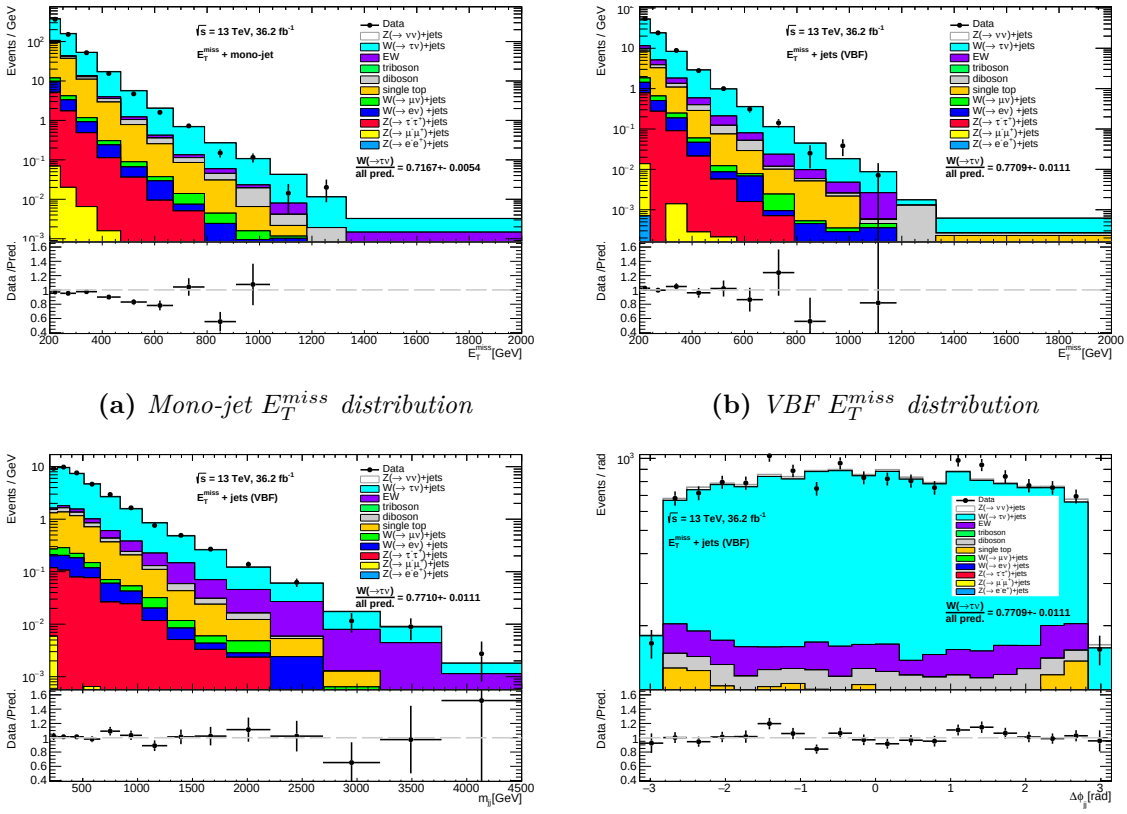


(c) *VBF m_{jj} distribution*



(d) *VBF $\Delta\phi_{jj}$ distribution*

Figure 3.1: *Agreement of data and Monte Carlo prediction for zero-lepton signal region.*



(a) *Mono-jet E_T^{miss} distribution*

(b) *VBF E_T^{miss} distribution*

(c) *VBF m_{jj} distribution*

(d) *VBF $\Delta\phi_{jj}$ distribution*

Figure 3.2: *Agreement of data and Monte Carlo prediction for single τ-lepton control region.*

3.1.2 Stability of the Control Region under different Selections

For the scale factor studies in the following sections, different selections are applied in order to observe the stability of the scale factor due to the selections. Therefore, the used selection criteria (section 2.5) are analyzed for the amount of $W \to \tau\nu_\tau$ and the agreement between the data and the Monte Carlo predictions. In figure 3.3 and 3.4 the W-boson transverse mass M_T and the τ-lepton transverse momentum p_T are displayed under the defined selections in the single τ-lepton control region.

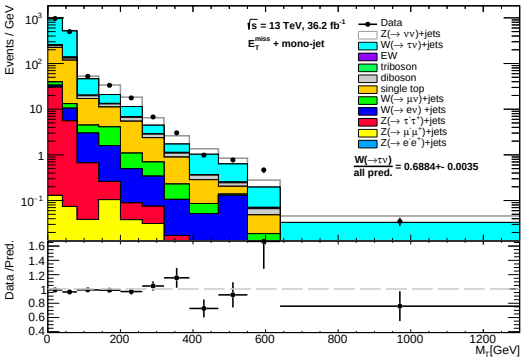
The M_T distribution of the mono-jet and VBF signature is shown for the standard, the unrestricted and the restricted selection in figure 3.3. $W \to \tau\nu_\tau$ has an amount of 68-72% in the mono-jet signature with an uncertainty of about 0.03-0.07% and an amount of 74-78% in the VBF signature with an uncertainty of about 0.07-1.5%. Following from that, the background contribution in the corresponding control region is quite small. In general, it can be noticed that the lowest amount of $W \to \tau\nu_\tau$ is found for the unrestricted selection and from there on the relative contribution increases by restricting the M_T spectrum. Since the full spectrum (fig.3.3a and 3.3b) also includes signals at the edges of the spectrum where less events are counted in the tail, it is reasonable to presume a smaller amount of $W \to \tau\nu_\tau$ within this range. The contribution of $W \to \tau\nu_\tau$ in the restricted M_T selection (fig.3.3e and 3.3f) does not change a lot comparing to the standard selection (fig.3.3c and 3.3d) because the range only gets tighter by 10 GeV on both edges, excluding only a few events.

In the full M_T selection both signatures have a good agreement for the data and the Monte Carlo simulation, deviating from one for higher M_T ranges. The VBF Monte Carlo prediction is underestimated for higher M_T ranges.

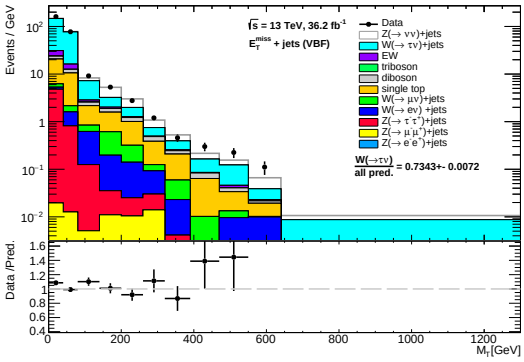
The standard M_T selection between 30-100 GeV shows a Monte Carlo simulation overestimation by 5-10% for most of the bins. For high M_T ranges the disagreement of the data to prediction events gets higher up to 40% for the VBF signature, simultaneously the uncertainties increase up to 15% and thus partially cover the deviation.

The restricted M_T selection between 40-90 GeV looks nearly as similar as the standard selection. Except of a few bins that are more lying within the range, the structure did not change.

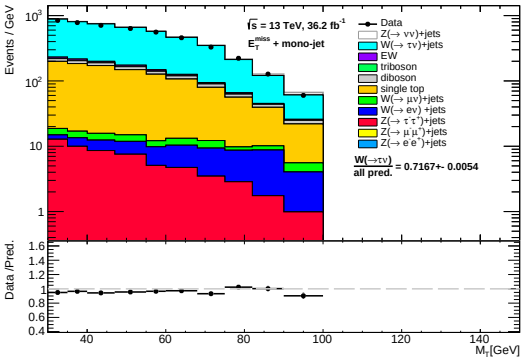
In figure 3.4 the τ-lepton momentum selection is represented. A general agreement of the data and the Monte Carlo prediction is recognizable. For the mono-jet signature (fig.3.4a and 3.4c) the Monte Carlo prediction are overestimated by up to 40%. The agreement in the VBF signature (fig.3.4b and 3.4d) varies strongly especially for high p_T. The statistical uncertainties are very large in this range and vary up to 50%. The deviations between the standard momentum selection and the restricted selections are small, since the spectrum was also constrained here only by 10 GeV on the lower edge.



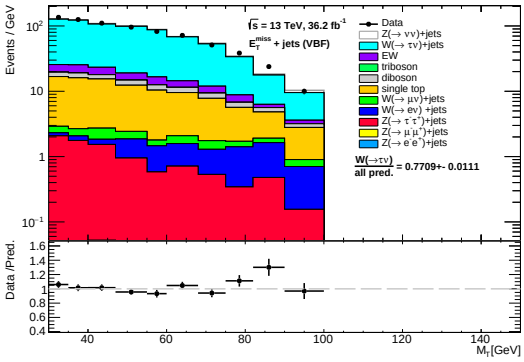
(a) *Mono-jet unrestricted M_T selection*



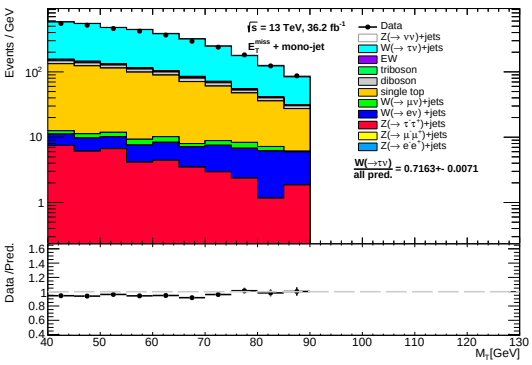
(b) *VBF unrestricted M_T selection*



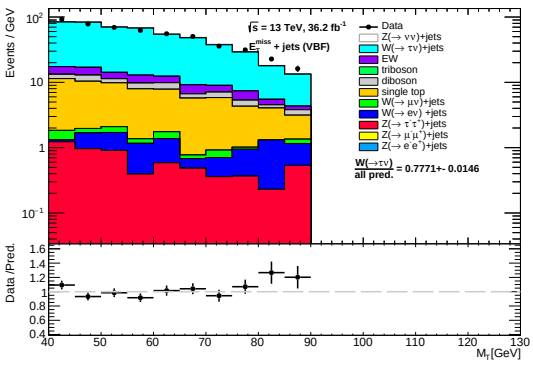
(c) *Mono-jet standard M_T selection*



(d) *VBF standard M_T selection*

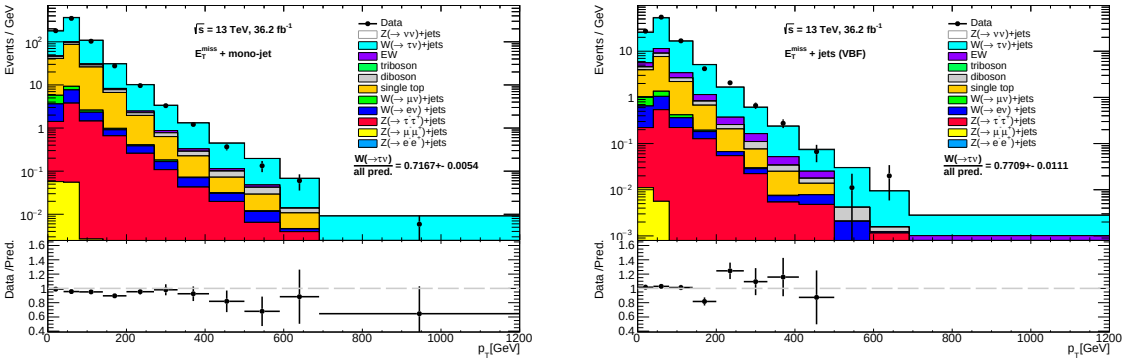


(e) *Mono-jet restricted M_T & p_T selection*



(f) *VBF restricted M_T & p_T selection*

Figure 3.3: *Agreement of data and Monte Carlo prediction of the W-boson M_T distribution for single τ-lepton control region.*



(a) *Mono-jet standard p_T selection*

(b) *VBF standard p_T selection*

(c) *Mono-jet restricted M_T & p_T selection*

(d) *VBF restricted M_T & p_T selection*

Figure 3.4: *Agreement of data and Monte Carlo prediction of the τ-lepton p_T distribution for single τ-lepton control region.*

3.2 Normalization of W $\rightarrow \tau\nu_\tau$ in the Signal Region

3.2.1 Scale Factor Definition

This thesis mainly attempts to determine the normalization of the $W \rightarrow \tau\nu_\tau$ background prediction in the signal region. This is done by scale factors, which will be calculated by means of the sum over all events in all bins of the data and the Monte Carlo prediction corresponding to the single τ-lepton control region. Afterwards, the scale factor will be applied to the number of $W \rightarrow \tau\nu_\tau$ predicted events in the signal region.

To achieve this, the $W \rightarrow \tau\nu_\tau$ Monte Carlo simulated events are compared with the events of the data subtracted by the Monte Carlo background, excluding the $W \rightarrow \tau\nu_\tau$ process. Therefore, the first step is to calculate this amount of data N^W_{data} as in the following:

$$N^W_{data} = N_{data} - N^{background}_{pred.}. \tag{3.9}$$

The corresponding uncertainty is calculated as

$$\Delta N^W_{data} = \sqrt{(\Delta N_{data})^2 - (\Delta N^{background}_{pred.})^2} \tag{3.10}$$

$$= \sqrt{(\sqrt{N_{data}})^2 - (\sigma^{background}_{pred.})^2} \tag{3.11}$$

$$= \sqrt{N_{data} - (\sigma^{background}_{pred.})^2}. \tag{3.12}$$

In the formulas above, N represents the number of events and the index *background* represents the number of background events in the control region. $\sigma^{background}_{pred.}$ describes the statistical uncertainty of the Monte Carlo background and is calculated as in equation (3.3). The scale factor (SF) is not calculated bin by bin but with the ratio of the total number of data events subtracted by the Monte Carlo background N^W_{data} and the total number of $W \rightarrow \tau\nu_\tau$ events $N^W_{pred.}$ as follows

$$SF = \frac{N^W_{data}}{N^W_{pred.}}. \tag{3.13}$$

The uncertainty of the scale factor σ_{SF} is determined with the error calculation in equation (3.2) and (3.3) as

$$\sigma_{SF} = SF * \sqrt{\frac{1}{N^W_{data}} + \Big[\frac{\sigma^W_{pred.}}{N^W_{pred.}}\Big]^2}. \tag{3.14}$$

3.2.2 Application of the Scale Factor

In order to determine the scale factor, the single τ-lepton event selection was required for data and all background processes. The observables of interest for data were corrected by subtracting all background processes using Monte Carlo simulation, not taking the $W \rightarrow \tau\nu_\tau$ process into account. In figure 3.5 the mono-jet E_T^{miss} distribution for the standard selection is shown for the corrected data and the $W \rightarrow \tau\nu_\tau$ simulation. The lower panel indicates the scale factor with its statistical uncertainty for each bin. If the amount of data and the Monte Carlo simulation do not show large discrepancies, the scale factor value should be around one. For the mono-jet signature in figure 3.5, the following scale factor was determined:

$$SF = 0.9422 \pm 0.0138. \tag{3.15}$$

Since the scale factor is less than one, this indicates an overestimation by the Monte Carlo prediction.

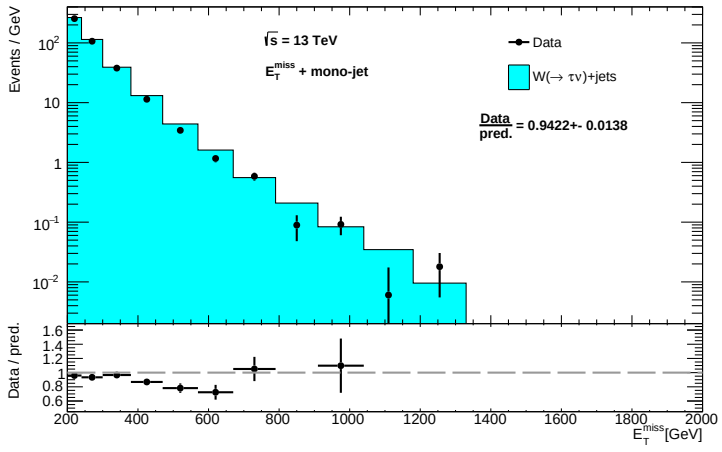


Figure 3.5: *Shown is the determination of the scale factor by comparing the data (black points) to the Monte Carlo prediction of $W \rightarrow \tau\nu_\tau$ (blue) for the mono-jet E_T^{miss} distribution for standard selection criteria in the single τ-lepton control region. The mean value of the scale factor is given.*

The calculated scale factor is applied to the signal region now. As the scale factor represents the disagreement of the data to Monte Carlo prediction corresponding to the single τ-lepton control region, the factor is used to correct the normalization of the $W \rightarrow \tau\nu_\tau$ events in the signal region. The number of these events will be multiplied bin by bin with the scale factor (SF) as

$$N_{pred.}^{SF} = SF * N_{pred.}^{W} \tag{3.16}$$

with the uncertainty

$$\sigma^{SF}_{pred.} = \sqrt{[\sigma_{SF} * N^{W}_{pred.}]^2 + [SF * \sigma^{W}_{pred.}]^2}, \tag{3.17}$$

leading to a correction of the ratio in equation (3.7). Applying the number of $W \rightarrow \tau\nu_\tau$ events with applied scale factor $N^{SF}_{pred.}$ on the ratio leads to

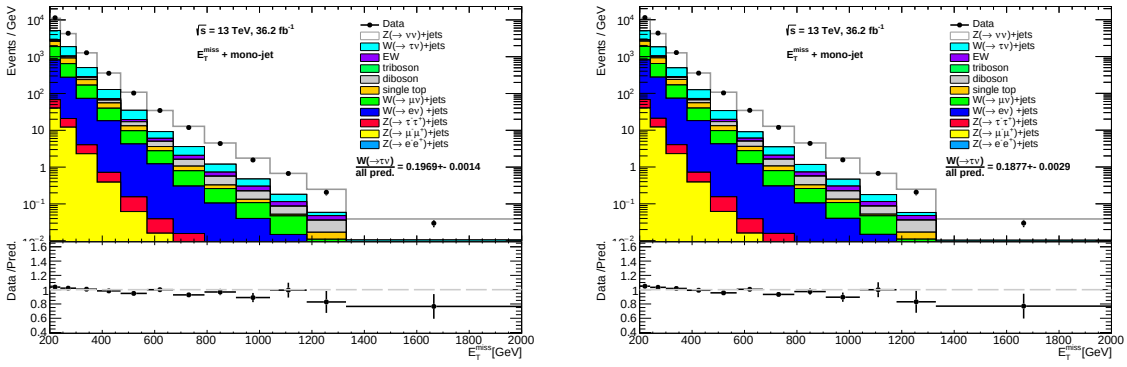
$$R^{corr.}_{W\rightarrow\tau\nu_\tau} = \frac{N^{SF}_{pred.}}{N^{all}_{pred.}}. \tag{3.18}$$

The uncertainty is calculated similar to equation (3.4) as

$$\Delta R^{corr.}_{W\rightarrow\tau\nu_\tau} = R^{corr.}_{W\rightarrow\tau\nu_\tau} * \sqrt{\left[\frac{\sigma^{SF}_{pred.}}{N^{SF}_{pred.}}\right]^2 + \left[\frac{\sigma^{all}_{pred.}}{N^{all}_{pred.}}\right]^2}. \tag{3.19}$$

Using the scale factor, presented in equation (3.15) for the $W \rightarrow \tau\nu_\tau$ normalization in the signal region, leads to smaller ratios. This is shown in figure 3.6. The mono-jet E_T^{miss} distributions in the signal region are displayed without applying the scale factor (3.6a) and taking the scale factor into account (3.6b). Without applying the scale factor, the total number of expected Monte Carlo events is overestimated by up to 20%.

This is visible in the ratio of the data to the Monte Carlo predictions in the lower panel of figure 3.6a.



(a) *No scale factor applied.* **(b)** *Scale factor applied.*

Figure 3.6: *Agreement of data and Monte Carlo prediction for the E_T^{miss} distribution. The mono-jet standard selection without and with applied scale factor is displayed in the signal region.*

The amount of $W \rightarrow \tau\nu_\tau$ events to be expected in the signal region lies at:

$$R_{W\rightarrow\tau\nu_\tau} = 0.1969 \pm 0.0014. \tag{3.20}$$

Making use of equation (3.18) and the scale factor derived in the control region, it is investigated how the relative contribution of $W \rightarrow \tau\nu_\tau$ changes. The results of this application are visible in figure 3.6b. The Monte Carlo predictions are still overestimated, however, the agreement of the data to the simulation shifts up around 1%. A confirmation is given by the corrected ratio:

$$R^{corr.}_{W\rightarrow\tau\nu_\tau} = 0.1877 \pm 0.0029. \tag{3.21}$$

It gets clear that the by a scale factor corrected $W \rightarrow \tau\nu_\tau$ ratio gets higher or lower depending on whether the scale factor is larger or smaller than one. The uncertainty of the ratio increased by a factor of two due to the larger uncertainty of the scale factor. Nevertheless, the comparison between the uncertainties of the ratio before and after the correction is not entirely fair. The $W \rightarrow \tau\nu_\tau$ normalization is determined in the control region along with its uncertainties and applied in the signal region. Before the application of the scale factor, the $W \rightarrow \tau\nu_\tau$ ratio in the signal region does not include uncertainties for the normalization of the Monte Carlo prediction, but the scale factor that is applied does, so the with the scale factor determined ratio has these uncertainties included, whereas the ratio without applied scale factor does not.

3.2.3 Comparison of Results

In the following, the scale factors of the mono-jet and VBF E_T^{miss} distribution will be determined and the modified selection criteria (section 2.5) will be included to see if and how the scale factor changes due to different selections. An ideal result would show that the scale factor is relatively stable due to the varying selections, as this allows to determine a suitable factor.

An overview of all scale factors for all relevent selection categories in the mono-jet and VBF signature is summarized in table 3.1. For each selection criteria the scale factor for the mono-jet and VBF E_T^{miss} distribution is calculated and will be used to determine a corrected ratio between the data and the $W \rightarrow \tau\nu_\tau$ Monte Carlo prediction in the signal region. The results from the table are shown visually in figure 3.7. There, the scale factor and the corrected ratio in the signal region are compared between the different selections.

Standard Selection

The results for the mono-jet E_T^{miss} distribution for the standard selection were already discussed in section 3.2.2. In this section, the standard selection for all observables in the VBF region will be discussed.

The scale factor is determined on the basis of the measured events between the data and the Monte Carlo simulation in the control region, as there exist the same number of events for every observable with a specific signature, the scale factor for the VBF E_T^{miss} distribution is similar to the scale factor of the VBF m_{jj} and $\Delta\phi_{jj}$ distribution. Therefore, the VBF E_T^{miss} distribution is analyzed and displayed in the following, supposing that the same can be assigned to the VBF m_{jj} and $\Delta\phi_{jj}$ distribution.

The scale factor for the VBF signature is around 7-8% larger than for the mono-jet signature. It is larger than one, thus the Monte Carlo predictions are underestimated. The scale factor has an uncertainty of 3.1%, which is around 2% higher than for the mono-jet signature.

The larger uncertainties can be explained with the number of events belonging to the $W \rightarrow \tau\nu_\tau$ process in the single τ-lepton control region. For the VBF signature less events can be counted in total compared to the mono-jet signature.

In order to confirm this, the agreement between the data and the $W \rightarrow \tau\nu_\tau$ Monte Carlo prediction in the control region is shown in Appendix A.1. The mono-jet and VBF E_T^{miss} distribution and the VBF m_{jj} and $\Delta\phi_{jj}$ distribution are displayed. The number of events for the data without Monte Carlo background and the number of $W \rightarrow \tau\nu_\tau$ events are given in the figures in Appendix A.1. The mono-jet signature has ten times more events than the VBF signature. The amount of data is smaller than the amount of Monte Carlo predictions in the mono-jet phase space and larger in the VBF phase space, that is responsible for the over- or underestimation of the Monte Carlo simulations. As mentioned, a similar number of events for the data and the $W \rightarrow \tau\nu_\tau$ process is recognizable for the VBF E_T^{miss}, m_{jj} and $\Delta\phi_{jj}$ distribution, since all events are sorted into the same phase space.

Due to the Monte Carlo prediction underestimation in the VBF signature the scale factor increases the corrected ratio of the $W \rightarrow \tau\nu_\tau$ process and its uncertainty in the signal region.

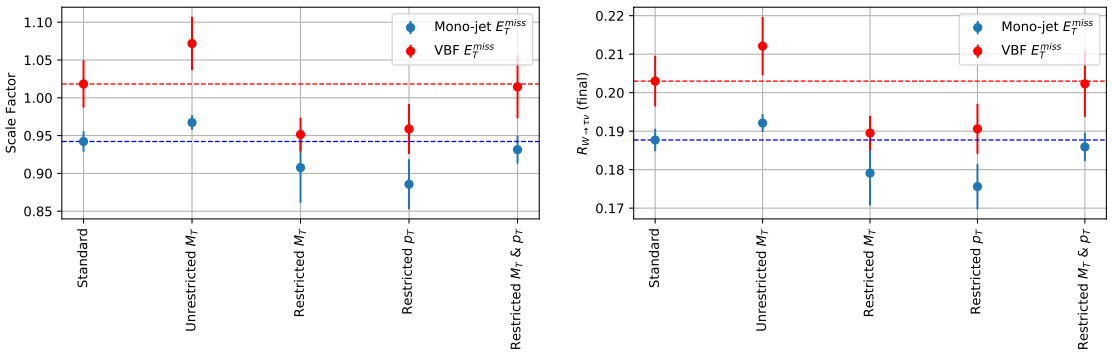
Modified Selection Criteria

The different results for the modified selection criteria are shown in figure 3.7. The scale factor (fig.3.7a) and the corrected $W \rightarrow \tau\nu_\tau$ amount (fig.3.7b) for the mono-jet and VBF E_T^{miss} distribution for all five selections are displayed. For both graphs, a horizontal line was added, depicting the nominal value of the scale factor and the corrected ratio using the standard selection.

The same course can be observed in the graphs. The scale factors change with different selections, which implies that they are not independent and deviate even

		Standard Selection	Unrestricted M_T Selection	Restricted M_T Selection	Restricted p_T Selection	Restricted p_T & M_T Selection
Mono-jet E_T^{miss}	SF	0.9422 ± 0.0138	0.9673 ± 0.0096	0.9076 ± 0.0466	0.8856 ± 0.0425	0.9314 ± 0.0184
	$R_{W\rightarrow\tau\nu}$ (initial)	0.1969 ± 0.0014	0.1973 ± 0.0013	0.1973 ± 0.0011	0.1938 ± 0.0011	0.1969 ± 0.0014
	$R_{W\rightarrow\tau\nu}$ (final)	0.1877 ± 0.0021	0.1921 ± 0.0021	0.1791 ± 0.0018	0.1756 ± 0.0020	0.1859 ± 0.0020
VBF E_T^{miss}	SF	1.0183 ± 0.0312	1.0718 ± 0.0223	0.9514 ± 0.0353	0.9588 ± 0.0331	1.0145 ± 0.0416
	$R_{W\rightarrow\tau\nu}$ (initial)	0.2000 ± 0.0017	0.2008 ± 0.0013	0.1973 ± 0.0016	0.1973 ± 0.0016	0.2000 ± 0.0017
	$R_{W\rightarrow\tau\nu}$ (final)	0.2030 ± 0.0066	0.2121 ±0.0076	0.1895 ± 0.00345	0.1906 ± 0.0065	0.2023 ± 0.0086

Table 3.1: *Overview of all scale factors (SF) for the E_T^{miss} mono-jet and VBF signature and the $W \rightarrow \tau\nu_\tau$ contribution in the signal region without (initial) and with (final) applied scale factor for different selection criteria.*



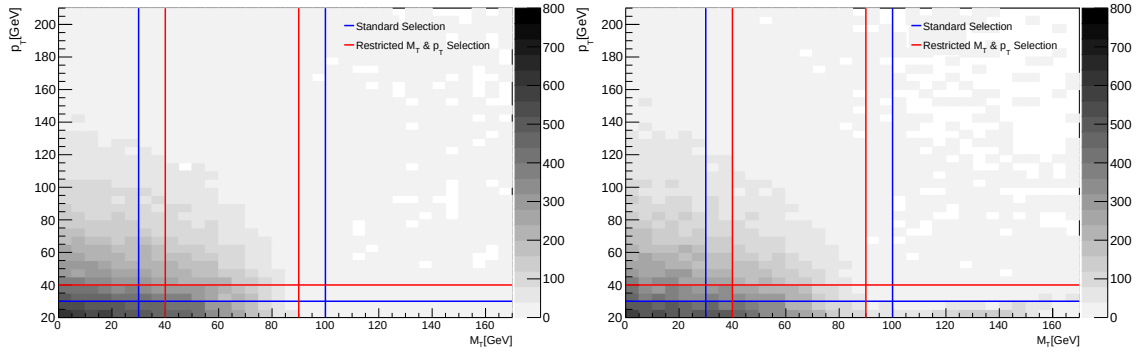
(a) *Scale Factor.* **(b)** *Corrected ratio* $R^{corr.}_{W\rightarrow\tau\nu_\tau}$.

Figure 3.7: *Results for the scale factor and the corrected* $W \rightarrow \tau\nu_\tau$ *amount listed for different selections. Displayed are the mono-jet and VBF* E_T^{miss} *distribution.*

more from each other than the statistical uncertainties would allow. For all selections the results for the VBF signature are higher than the results for the mono-jet signature. Moreover, the uncertainties of the VBF signature are usually larger than in the mono-jet signature, since the sum of all events in the VBF region is smaller than in the mono-jet region. The corrected $W \rightarrow \tau\nu_\tau$ amount contributes around 17-19% to the mono-jet signature and 19-21% to the VBF signature of the signal region, the uncertainties have become larger by 0.1% and 0.3% respectively.

It is visible that for restricted selections the values of the scale factor and the corrected ratio become much lower compared to the standard selection. For the restriction of both parameters M_T and p_T the results change only insignificantly by a factor of about 1%. This is further reinforced by restricting the mass and momentum. By restricting the bounds of the standard selection in the single τ-lepton control region, only fewer events are selected. The scale factor becomes smaller for the mono-jet and VBF signature, which means that by restricting the M_T and p_T range the number of measured data events decreases more than the number of Monte Carlo simulated events with the result that the Monte Carlo prediction is even more overestimated.

The reason for the larger difference between the restricted M_T & p_T selection and the restriction of a single parameter can be explained by figure 3.8. The dependence of the transverse mass of the W-boson on the transverse momentum of the τ-lepton is presented. The z axis corresponds to the number of events for the $W \rightarrow \tau\nu_\tau$ process in figure 3.8a and for the data in figure 3.8b, whereby the data was subtracted by the number of Monte Carlo background events in the control region. The white spots in the figures are points where there are no events that satisfy the kinematic requirement in parameter space. The most events are found in the low energy range and become increasingly less with higher momentum and higher mass. The constraints of the standard selection and the M_T & p_T restricted selection have been added in the figure. The blue frame in the middle includes all events that pass the standard selection and the red frame in the middle includes the events that are part



(a) *$W \rightarrow \tau\nu_\tau$ Monte Carlo prediction distribution*

(b) *$W \rightarrow \tau\nu_\tau$ data distribution*

Figure 3.8: *The number of events displayed for mono-jet τ-lepton momentum p_T as a function of W-boson mass M_T in the single τ-lepton control region. Standard Selection and restricted M_T & p_T selection is applied.*

of the restricted selection of both parameters. If only one of the two parameters is restricted, either the horizontal or the vertical red line will be left out in the figure.

It can be seen that the population of events in the lower left corner is excluded from the restricted selection but is still included in the standard selection. Especially in this part, a lot of data is covered which is lost by the restriction. While the restriction of both parameters excludes the same amount of data and $W \rightarrow \tau\nu_\tau$ Monte Carlo prediction events, the restriction of one parameter at a time loses some of the events corresponding to the data, which is why there are more Monte Carlo simulation events, following consequently in a lower scale factor. Although only the mono-jet signature is displayed, the event distribution as in figure 3.8b looks similar for the VBF signature (Appendix A.2), therefore the same statements can be made.

The scale factor for the VBF signature decreases strongly for the restricted M_T selection by about 7% compared to the standard selection and by 6% for the restricted p_T selection. The scale factor for the mono-jet signature decreases by 4% compared to the standard selection for the restricted M_T selection and by 6% for the restricted p_T selection. These contributions suggest that restricting the W-boson mass has a higher influence on the scale factor of the VBF signature, whereas restricting the τ-lepton momentum has a higher influence on the scale factor of the mono-jet signature. This can be traced back to kinematic differences in both signatures or just to statistical variances.

The results for the scale factor and the corrected $W \rightarrow \tau\nu_\tau$ ratio for the unrestricted M_T mass are larger than for other selections. The scale factor value increased because events even for high M_T ranges are included and a prediction at high energy is difficult, resulting in less predictions are taken into account. Besides, in figure 3.8b a tail of data events can be recognized for higher M_T ranges confirming a higher amount of data events in this area than an amount of Monte Carlo simulated events. The uncertainties for the full spectrum are much lower than for the other selections. If the transverse mass is not constrained, many more events are included from lower energy levels. This increased amount of events leads to a

reduction of the uncertainty.

3.2.4 Applicability

Due to the fact, that the scale factor is determined by the relation of data to Monte Carlo prediction in the single τ-lepton control region, it needs to be checked whether the scale factor can be applied on the signal region.

In order to compare the amount of the $W \rightarrow \tau\nu_\tau$ Monte Carlo prediction in the signal region to the amount in the control region, it is necessary to compare the distributions with τ-lepton influence, namely the W-boson M_T and the τ-lepton p_T distribution, as they are the only deviating parameters. Additionally, the total number of events needs to be the same in both region, which is why the number of events for each bin is normed to one. Since there are no leptons detected in the signal region, the Monte Carlo predictions will be analyzed on particle-level with reconstruction cuts.

On particle-level every decay process is visible in the sense that the final-state is known. That means, that one can recognize all τ-leptons, which will be misidentified and therefore not rejected in the signal region. The reconstruction cut makes sure that the τ-lepton will be assigned to the signal region, although the final state is known and would be otherwise rejected.

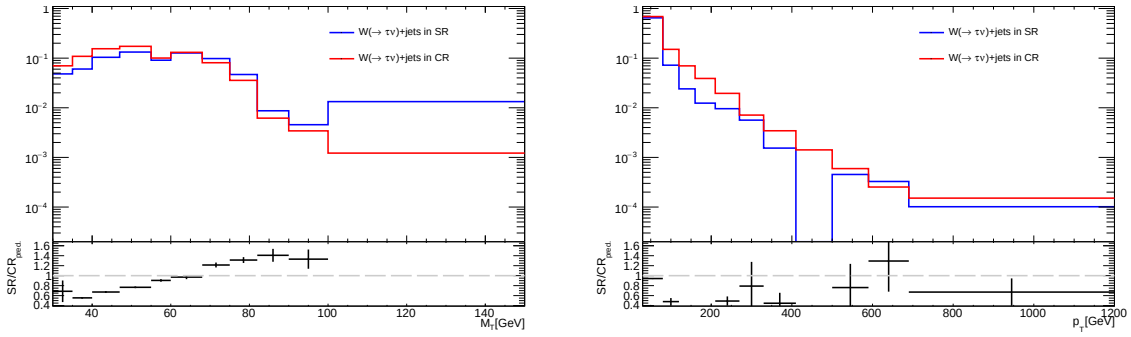
In figure 3.9 the mono-jet signature for the M_T (fig.3.9a) and p_T (fig.3.9b) distribution is shown for the standard selection. As the particle-level estimations are a simulation, only the Monte Carlo prediction samples of the $W \rightarrow \tau\nu_\tau$ process are used. The distribution in the signal region, colored in blue, is displayed next to the distribution in the control region, colored in red. The ratio of the normed number of Monte Carlo predicted events in the signal region regarding to the control region $R_{SR/CR}$ is shown bin by bin in the lower panel of the graphs and is calculated as follows

$$R_{SR/CR} = \frac{N^W_{SR}}{N^W_{CR}}. \tag{3.22}$$

N^W indicates the normed number of simulated $W \rightarrow \tau\nu_\tau$ events in the signal region SR or in the control region CR. The uncertainty of the ratio $\Delta R_{SR/CR}$ is determined by

$$\Delta R_{SR/CR} = R_{SR/CR} * \sqrt{\left[\frac{\sigma^W_{SR}}{N^W_{SR}}\right]^2 + \left[\frac{\sigma^W_{SR}}{N^W_{CR}}\right]^2}. \tag{3.23}$$

The ratio of both regions shifts upwards with increasing energy in the M_T distribution and shifts up- and downwards in the p_T distribution. The eventless bin in the signal region in figure 3.9b is a negative-weight event and shall be ignored. The normed number of $W \rightarrow \tau\nu_\tau$ events in the signal region only differs slightly from the numbers of events in the control region, which in general approves the applicability of the scale factor.



(a) *W-boson M_T distribution* **(b)** *τ-lepton p_T distribution*

Figure 3.9: *Comparison of the normed number of Monte Carlo predicted $W \rightarrow \tau\nu_\tau$ events in the signal region (SR) to the control region (CR) in the mono-jet signature on particle-level with reconstruction cut. Standard selection is applied.*

3.3 Systematic Uncertainties

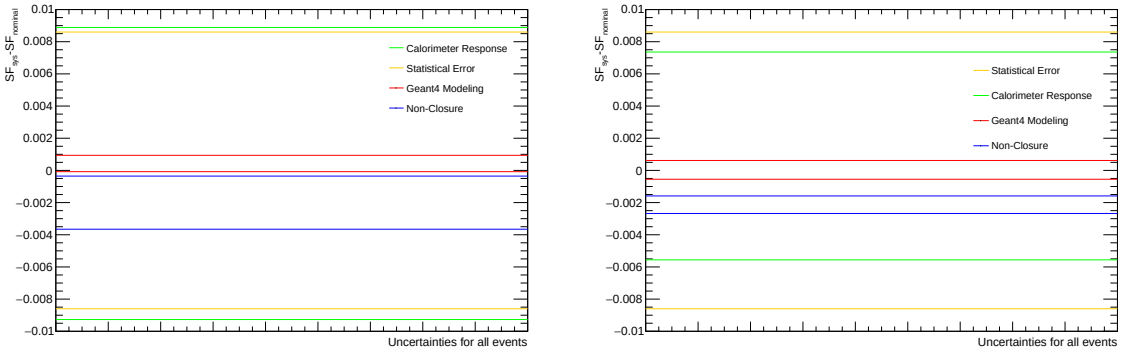
After the statistical uncertainties have been included, the systematic uncertainties are analyzed and compared to the statistical uncertainties in the following. The systematic uncertainties used all stem from uncertainties in the reconstruction process of τ-leptons and include one-prong and three-prong decays. Three different systematic uncertainties are studied. More detailed information about the determination of these uncertainties can be found in [14].

The *Geant4 modeling* describes uncertainties due to the hadronic shower shape modeling. Hadrons in jets are measured as charged-particle tracks or as showers of particles in the calorimeter. For reconstruction reasons, these showers have to be analyzed by their shape. The shower shape model is correlated to the energy density in a shower, so that a variation of the shower shape model influences the measured energy distribution in a cluster. The τ-lepton energy reconstruction is based on the *local hadron calibration weight (LCW)* which changes due to a different cluster energy. In order to determine the magnitude of the uncertainty, different shower shape models were compared to each other.

The *calorimeter response* uncertainty describes the deviations from the hadronic τ-lepton energy scale calibration. The uncertainty is derived from the convolution of the hadronically decaying τ-lepton products, where the events are analyzed without any correction to allow a pure calorimeter response uncertainty.

The last included systematic uncertainty is the *non-closure* of the *LCW* calibration method. This uncertainty is relevant in the process of determining the calibration constant, which is necessary for calculating the calibrated energy of the τ-lepton.

The contribution of the τ-lepton systematic uncertainties compared to the contribution of the statistical uncertainties is visible in figure 3.10. The systematic and statistical uncertainty is shown for the VBF and mono-jet signature. For each uncertainty a 1σ-deviation up- and downwards was determined. The scale factor of



(a) *Mono-jet signature.* **(b)** *VBF signature.*

Figure 3.10: *Differences between the systematic scale factor and the nominal scale factor in the single τ-lepton control region displayed for all events in the corresponding signature.*

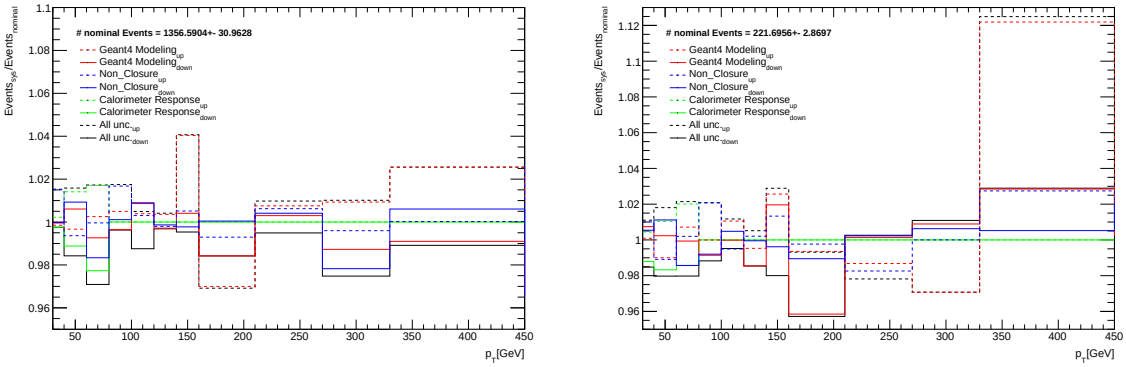
the systematic data sets were calculated and subtracted by the scale factor of the nominal data set. The statistical uncertainty of the nominal scale factor, calculated as in equation (3.14), is also shown in the graphs.

For the mono-jet signature, figure 3.10a exemplifies that the statistical uncertainty does not differ remarkably from the calorimeter response, both of them have a percentage of about 0.9% up- and downwards. In contrast, the other systematic uncertainties are smaller. The Geant4 modeling deviates 0.1% upwards, whereas the uncertainty coupled to the non-closure deviates 0.35% downwards.

In figure 3.10b the statistical uncertainty deviates considerably more (2.39%) with respect to the calorimeter response, which is due to the high uncertainty of the scale factor for the VBF signature. The uncertainty coupled to the calorimeter response has a 0.8% upwards deviation and a 0.6% downwards deviation. The other systematic uncertainties are nearly similar to the mono-jet signature, they only deviate up to 0.3%.

The reason for the high uncertainty value of the calorimeter response can be explained by the calibrated energy resolution. Due to the desired linear relationship of the measured energy to the actual energy of a shower, the resolution ratio should remain constant as a function of energy. However, the resolution only remains approximately constant for lower energy ranges and increases exponentially for higher energy ranges. Since the calorimeter response affects the τ-lepton energy scale, the uncertainty is correspondingly large for high-energy τ-leptons.

Moreover, in figure 3.11 the number of events for each systematic uncertainty has been divided by the number of nominal events as a function of the τ-lepton momentum. The up and down fluctuation is marked and the sum of all systematic uncertainties were added and displayed. The calorimeter response only fills the range of about 30-80 GeV, indicating that the nominal and systematical events have a good agreement for high momentum ranges. This is due to the fact that the calorimeter response depends on the p_T distribution, so that more accurate results can be measured for high momentum ranges. The Geant4 modeling and non-closure fluctuate strongly for high momentum scales. It gets more difficult to simulate τ-

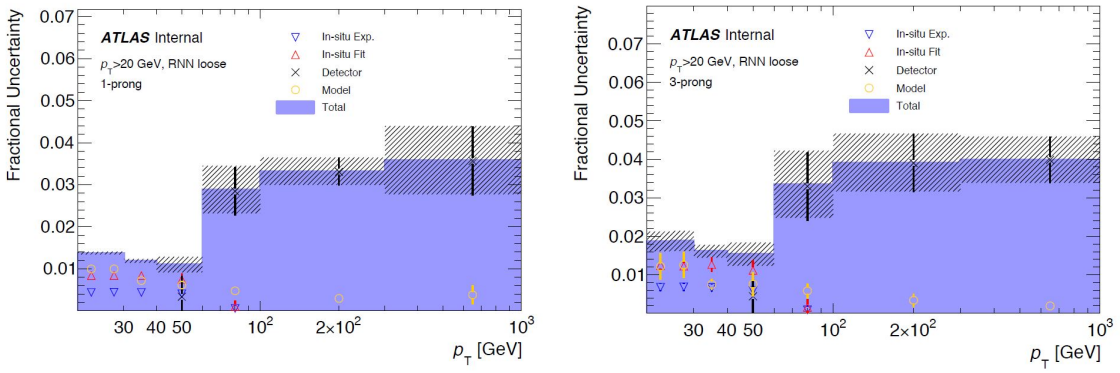


(a) *Mono-jet signature.* **(b)** *VBF signature.*

Figure 3.11: *Deviations in the number of Monte Carlo predicted events for systematic events to nominal events for the τ-lepton p_T distribution in the single τ-lepton control region. In addition, the sum for all up- or downwards deviating uncertainties were taken (black).*

leptons with higher momentum, therefore less events are counted in total and the statistics vary greatly.

This can be compared to the graphs in figure 3.12. The systematic uncertainties for the τ-lepton energy scale were determined for one-prong and three-prong decays, the figures were taken from [15]. The results in 3.12 correlate to the results determined in figure 3.10. The combination of in-situ experimental and in-situ fit corresponds to the calorimeter response, whereas the uncertainty depicted by model corresponds to the non-closure uncertainty. The uncertainties in this thesis were determined for one- and three-prong decays inclusively and were not separated. For a lower energy scale, the uncertainties are much higher and decrease with increasing energy. The calorimeter response also settles at about 0.9%, only the non-closure values are slightly higher than the ones in 3.10.



(a) *One-prong τ-lepton decay.* **(b)** *Three-prong τ-lepton decay.*

Figure 3.12: *τ-lepton energy scale uncertainties in GeV for one-prong and three-prong τ decay. In-stitu corresponds to calorimeter response, whereas model corresponds to non-closure [15].*

In conclusion, the statistical uncertainty of the scale factor is mostly larger than the systematic uncertainty, which is shown in figure 3.10. Although the calorimeter response is a bit larger than the statistical uncertainty for the mono-jet signature in figure 3.10a, neither the calorimeter response nor the statistical uncertainty are in general sufficient to compensate the discrepancy of the scale factor values for different selections, which means that the scale factors of the individual selections differ significantly from each other.

4 Conclusion

In order to search for physics beyond the Standard Model, it is necessary to precisely control the Standard Model backgrounds in any signal region related to dark matter production. Hence, this thesis has analyzed the decay of the W-boson into a neutrino and a τ-lepton, as well as the corresponding single τ-lepton control region. The background contribution of the $W \to \tau\nu_\tau$ process in the signal region was found to be relatively high (20%) and outweighs the other background processes.

The normalization of the Monte Carlo $W \to \tau\nu_\tau$ process in the signal region was corrected by means of a scale factor derived in the control region. The application of this scale factor in the signal region using standard selections led to a decrease of the $W \to \tau\nu_\tau$ amount in the mono-jet region and to an increase in the VBF region, due to an over- and underestimation from the Monte Carlo simulation.

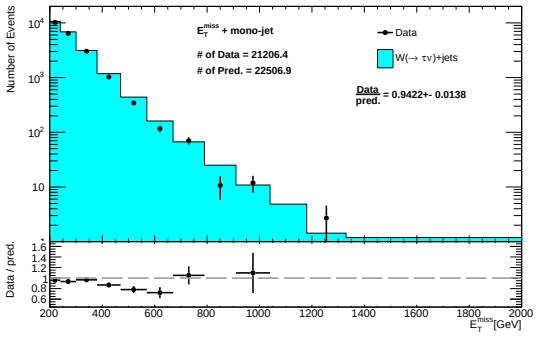
In addition, different modified selection criteria were applied and it was observed that the scale factors determined by constraining or extending the transverse W-boson mass or the transverse τ-lepton momentum ranges vary from the result with the standard selection. The statistical uncertainties cannot cover the discrepancy of the scale factor values for different selections. The determination of a suitable scale factor is therefore difficult.

Moreover, various scale factors calculated with the systematic uncertainties from the τ-lepton reconstruction process were compared with the nominal scale factor. In principal, the statistical uncertainty is larger than the systematic uncertainties, except for the mono-jet signature where the calorimeter response uncertainty slightly outweighs the statistical uncertainty. Still, these uncertainties are not large enough to cover the discrepancy of the scale factor results to the standard selection.

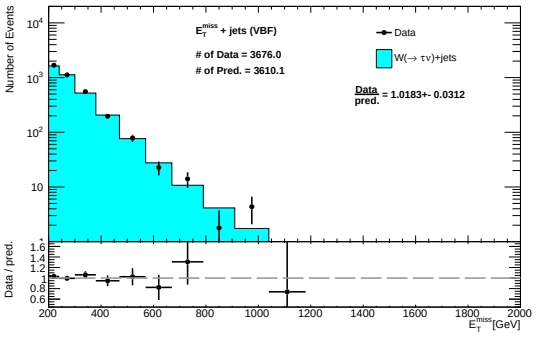
Further possibilities exist to correct the background contribution in the signal region. Scale factors could be determined in control regions from different Monte Carlo background processes and applied to not only the signal region, but to the other control regions. When the normalization of the background processes in those control regions is corrected, a more accurate scale factor would be determinable, which could lead to an exact normalization of those background processes in the signal region.

A Scale Factor Studies

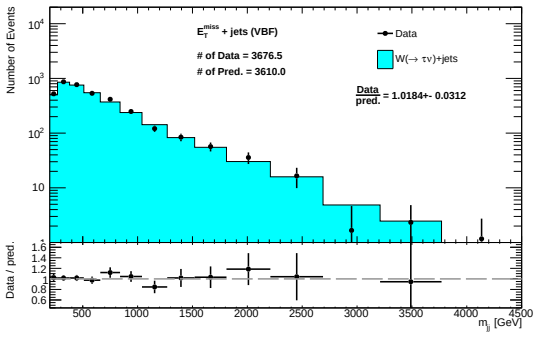
A.1 Determination of the Scale Factor



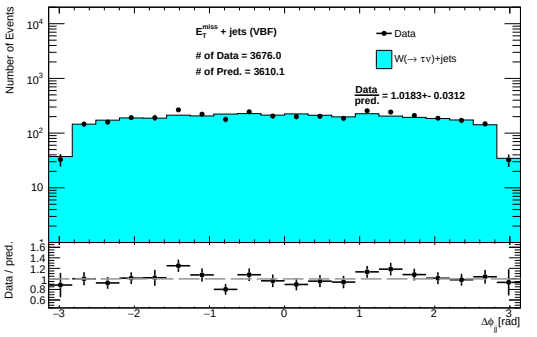
(a) *Mono-jet E_T^{miss} distribution*



(b) *VBF E_T^{miss} distribution*



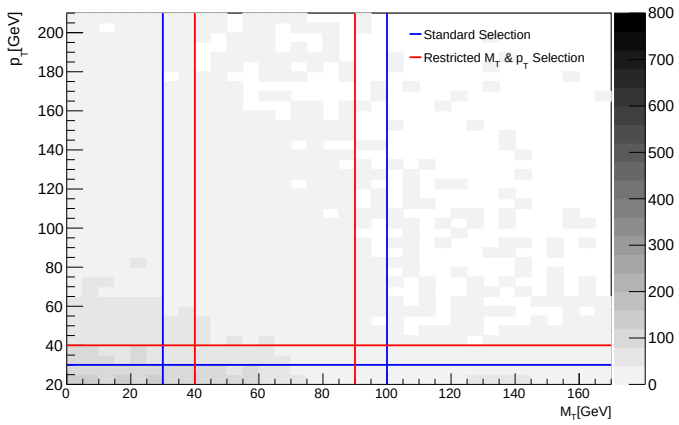
(c) *VBF m_{jj} distribution*



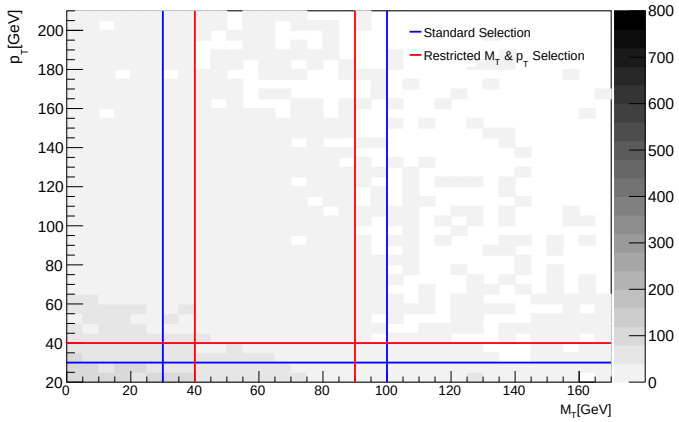
(d) *VBF $\Delta\phi_{jj}$ distribution*

Figure .1: *Agreement of Data to Monte Carlo prediction for $W \to \tau\nu_\tau$ processes in the single τ-lepton control region. The amount of data and Monte Carlo prediction without background contribution is displayed. The total number of events of the data and Monte Carlo prediction is given for each distribution.*

A.2 M_T and p_T dependence for the VBF signature



(a) *$W \to \tau\nu_\tau$ Monte Carlo prediction distribution*



(b) *$W \to \tau\nu_\tau$ data distribution*

Figure .2: *The number of events displayed for VBF τ-lepton momentum p_T as a function of W-boson mass M_T in the single τ-lepton control region. Standard Selection and restricted M_T & p_T selection is applied.*

Bibliography

[1] Mark Thomson. *Modern Particle Physics*. Cambridge University Press, 2013.

[2] A.B. Arbuzov. *Quantum Field Theory and the Electroweak Standard Model*. 2018. arXiv: 1801.05670v1 [hep-ph].

[3] J. Pequenao. *Computer generated image of the whole ATLAS detector*. 2008. URL `https://cds.cern.ch/record/1095924`.

[4] ATLAS Collaboration. *The ATLAS Experiment at the CERN Large Hadron Collider*. 2008. ISSN JINST 3 S08003.

[5] Thomas Malte Spieker. *Detector Corrected Search for Dark Matter in Monojet and Vector Boson Fusion Topologies with the ATLAS Detector*. PhD thesis, Universität Heidelberg, 2019.

[6] ATLAS Collaboration. *Measurement of detector-corrected observables sensitive to the anomalous production of events with jets and large missing transverse momentum in pp collisions at $\sqrt{s}$=13 TeV using the ATLAS detector*. 2017. ISSN 1434-6052.

[7] Alec Aivazis. *Draw Feynman Diagram Online*. URL `https://feynman.aivazis.com/`. Last access: 04.01.21.

[8] Stanislav Suchek. *Search for Dark Matter produced in association with hadronically decaying Standard Model bosons using 36 fb^{-1} of pp collisions at the ATLAS detector*. PhD thesis, Universität Heidelberg, 2018.

[9] ATLAS Collaboration. *Measurement of detector-corrected observables sensitive to the anomalous production of events with jets and large missing transverse momentum in pp collisions at $\sqrt{s}$=13 TeV using the ATLAS detector*. 2017. doi: 10.1140/epjc/s10052-017-5315-6.

[10] Irene Vichou. *Performance of Missing Transverse Momentum Reconstruction in ATLAS with Proton-Proton Collisions at $\sqrt{s}$ = 7 TeV*. 2012. doi: 10.1088/1742-6596/404/1/012011.

[11] Stefania Xella. *Physics object reconstruction in the ATLAS experiment*. 2013. ISSN ATL-GEN-PROC-2013-001.

[12] G. Azuelos et al. *Electroweak Physics*. 2000. arXiv: hep-ph/0003275.

[13] ATLAS Collaboration. *Measurement of the electroweak production of dijets in association with a Z-boson and distributions sensitive to vector boson fusion in proton-proton collisions at $\sqrt{s}$ = 8 TeV using the ATLAS detector*. 2014. arXiv: 1401.7610v2 [hep-ex].

[14] ATLAS Collaboration. *Determination of the tau energy scale and the associated systematic uncertainty in proton-proton collisions at sqrts= 7 TeV with the ATLAS detector at the LHC in 2011*. 2012. ISSN ATLAS-CONF-2012-054.

[15] ATLAS Collaboration. *Tau Recommendations Summer 2019*. 2019. URL `https://twiki.cern.ch/twiki/pub/AtlasProtected/TauRecommendationsR21/ValidationPlots-2019-summer.pdf`.

Acknowledgements

I would like to thank Prof. Dr. Monica Dunford for giving me the opportunity to write my bachelor thesis in the ATLAS group and for supervising me at least once a week, although her schedule is quite packed.
Also I want to thank my advisor Martin Klassen, whom I could bombard daily with questions and who could always help me out with them.
I'm also very grateful for everyone in this group, who supported me by correcting a chapter in my thesis. Thanks again, Sebastian, Fer, Falk and Philipp.
At last I want to thank my family and co-family for being there and keeping me company during the lonely Corona times.

Erklärung

Ich versichere, dass ich diese Arbeit selbstständig verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Heidelberg, den 11.02.2021 ..... Katharina Bews .....