Masterthesis

Search for $H \rightarrow \tau\tau$ Decays in the Lepton-Hadron Final State using Multivariate Techniques in Proton-Proton Collisions at $\sqrt{s} = 13\,\mathrm{TeV}$ with the ATLAS Detector at the LHC

Frank Sauerburger

Albert-Ludwigs-Universität Freiburg

Masterthesis

Search for $H \rightarrow \tau\tau$ Decays in the Lepton-Hadron Final State using Multivariate Techniques in Proton-Proton Collisions at $\sqrt{s} = 13\,\mathrm{TeV}$ with the ATLAS Detector at the LHC

submitted by
Frank Sauerburger

Supervisor
Prof. Dr. Karl Jakobs

March 17, 2017

Albert-Ludwigs-Universität Freiburg

Version: final-cds-1-gd056416
Date: June 29, 2017
Author: Frank Sauerburger
Physikalisches Institut
Albert-Ludwigs-Universität Freiburg

Erklärung

Hiermit versichere ich, die eingereichte Masterarbeit selbständig verfasst und keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt zu haben. Wörtlich oder sinngemäß aus anderen Werken übernommene Inhalte wurden entsprechend den anerkannten Regeln wissenschaftlichen Arbeitens (lege artis) kenntlich gemacht. Ich erkläre weiterhin, dass die eingereichte Masterarbeit weder vollständig noch in wesentlichen Teilen Gegenstand eines anderen Prüfungsverfahrens war oder ist.

Ort, Datum .... Unterschrift ....

Abstract

Abstract
A multivariate analysis (MVA) using machine learning techniques to study the Standard Model decay of the Higgs boson to two τ leptons ($H \to \tau^+\tau^-$) is presented. The analysis focuses on the decay channel, in which one τ decays leptonically and the other τ decays hadronically. The background is estimated with a combination of Monte Carlo simulation and data-driven methods. A boosted decision tree (BDT) is trained on the background model and employed to classify events into background and signal to enhance the sensitivity of the analysis. The results of the analysis are compared to a cut-based analysis. The analysis is performed with a dataset of proton-proton collisions at a center-of-mass energy $\sqrt{s} = 13\,\text{TeV}$ taken with ATLAS detector at the LHC during Run 2.

Zusammenfassung
Eine multivariate Analyse (MVA) des Standardmodelprozesses des Zerfalls des Higgsteilchens in zwei τ Leptonen ($H \to \tau^+\tau^-$) wird präsentiert. Die Analyse konzentriert sich auf den Zerfallskanal, indem ein τ leptonisch und ein τ hadronisch zerfällt. Der Untergrund wird durch eine Kombination von Monte Carlo Simulationen und datengestützten Methoden abgeschätzt. Ein Boosted Decision Tree (BDT) wird auf dem Untergrundmodel trainiert und verwendet, um Ereignisse als Untergrund oder als Signal zu klassifizieren, und um dadurch die Sensitivität der Analyse zu erhöhen. Die Ergebnisse der Analyse werden mit einer cut-basierten Analyse verglichen. Die Analyse verwendet Daten aus Proton-Proton Kollisionen bei einer Schwerpunktsenergie von $\sqrt{s} = 13\,\text{TeV}$, die vom ATLAS-Detektor am LHC während Run 2 aufgezeichnet wurden.

Preface

The present document constitutes my master thesis and summarizes the results of my work during the last year.

The main focus of this thesis is a multivariate analysis (MVA) to search for Higgs particle decays to two τ leptons. Most of the effort was spent on the optimization of the multivariate analysis. This includes the usage of an elaborate cross validation scheme and a private Monte Carlo production to increase the training statistics. The goal of this optimization is to enhance the sensitivity of the multivariate analysis. As a comparison a cut-based analysis is also presented. The final results from a likelihood fit for the multivariate analysis and cut-based analysis are compared.

The first chapter introduces the topic of $H \to \tau\tau$ interactions and shows the importance of this decay channel. Chapter 2 summarizes the Standard Model and shows how the Brout–Englert–Higgs mechanism is used to generate the masses of fermions, especially τ leptons. Chapter 3 outlines the phenomenology at hadron colliders and the physics of the Higgs boson. The following chapter, Chap. 4, describes the Large Hadron Collider and the ATLAS experiment, which is used to collect the data for this thesis. The first four chapters give an introduction to the basics of hadron colliders and Higgs analyses with the ATLAS detector. Readers already familiar with this type of analysis can skip these chapters and start with the first analysis specific chapter, Chap. 5.

Chapter 5 outlines the analysis strategies of the multivariate analysis and the cut-based analysis. There the background model of the analyses and the event selection is defined. Chapter 6 constitutes the main chapter of this thesis. It describes the machine learning techniques used for the MVA and shows the procedure used to optimize it. Finally, Chap. 7 describes a likelihood fit and shows its results. The two analysis strategies are compared in terms of their sensitivity. The last chapter, Chap. 8, summarizes the conclusions of this thesis and outlines future improvements.

The submission of this master thesis is a milestone for my personal life and my experience with the physics community. This seems to be the correct moment to thank all the people that helped me during the last year. Certainly this work would not have been possible without your help.

First of all I would like to thank Karl Jakobs for his support and for giving me the opportunity to work on such an interesting topic. I would also like to thank

Minoru Hirose, Karsten Köneke and Lei Zhang for guiding me and giving me valuable advice. Thank you for sharing your experience. Furthermore I would like to thank the whole HLeptons analysis group without whom this analysis would not have been possible. Especially I would like to thank Clara Nellist who also helped me out if had I problems with the English language. This is also the time to thank all the group members in Freiburg. Without you the last year would not have been such a great experience. Thank you for creating this wonderful work environment.

Frank Sauerburger March 2017, Freiburg

Notation

The following conventions and notations are used throughout this thesis. Velocities are measured as fractions of the speed of light ($c = 1$), while angular momenta are measured in multiples of the reduced Planck constant ($\hbar = 1$). The Einstein summation rule is used, which means a summation over an index is implied, if the index appears as an upper and low index in a summand, *e.g.*, $\gamma^\mu \, \partial_\mu \to \sum_\mu \gamma^\mu \, \partial_\mu$. The term *lepton* is reserved for charged leptons from the first and second generation (electrons e, and muons μ), with the exception of chapter 2 where lepton also refers to neutrinos and tau leptons.

In the analysis specific chapters, lepton or leptonically decaying tau (or simply lep) refers to the charged lepton originating from a leptonically decaying tau, *i.e.*, $\tau^- \to \ell^- \bar{\nu}_\ell \nu_\tau$. The visible products of a hadronically decaying tau, *i.e.*, $\tau^- \to q\bar{q}' \nu_\tau$, are referred to as a hadronically decaying tau, a tau or τ_had (or simply had).

The term *leading* (*sub-leading*) particle refers to the decay product with the largest (second largest) transverse momentum. Various kinematic variables can be expressed or calculated for different particles. The particle under study is mentioned in the superscript of the kinematic variable, this means $p_\mathrm{T}^{j_0}$ refers to the transverse momentum p_T of the leading jet j_0. Kinematic variables, which depend on two particles are written with both particles as its superscript, this means $\Delta\eta^\mathrm{lep\,had}$ is the difference in pseudorapidity of the hadronically decaying tau (had) and the leptonically decaying tau (lep). Truth quantities, *i.e.*, physical quantities not changed or smeared out by detector effects, are annotated with a hat, so for example $\hat{p}_\mathrm{T}^\mathrm{lep}$ denotes the truth transverse momentum of the lepton.

Bold face variables are three-component objects in the context of Quantum Field Theory and multi-dimensional vectors in the context of machine learning and the likelihood fit. Throughout the thesis $\log(x)$ refers to the natural logarithm of x.

Contents

CHAPTER 1

Introduction

In 2012 the ATLAS (A Toroidal LHC ApparatuS) [1] and CMS (Compact Muon Solenoid) [2] collaborations announced the discovery of a new particle [3, 4] at the Large Hadron Collider (LHC) [5]. The discovered particle is compatible with the Higgs boson predicted by the Standard Model of particle physics. The Higgs boson, proposed in 1964 by Higgs [6, 7], Brout and Englert [8], Guralnik, Hagen and Kibble [9] is an important ingredient in the Standard Model. The Higgs boson solves several problems of the Standard Model in an easy and clean way. To only mention some aspects, the Higgs boson is necessary to restore unitarity and prevent the prediction of divergent event rates in WW scattering [10]. Another important implication of the Higgs boson and the Brout–Englert–Higgs mechanism (BEH) is that the weak gauge bosons can acquire mass. Without the BEH mechanism, massive gauge bosons are forbidden by the Standard Model since they would violate important properties of the model. In 1983 the massive gauge bosons W and Z were discovered experimentally [11, 12]. At that time these two bosons were the heaviest observed elementary particles, which contradicts the requirement of massless gauge bosons in the standard model without the BEH mechanism.

Besides being responsible for the masses of gauge bosons, the BEH mechanism can also be employed to generate the masses of fermions by introducing a Yukawa coupling of the fermions to the Higgs field. Such a coupling implies that the Higgs particle can decay into pairs of fermions. These couplings have been measured for tau (τ) leptons and bottom (b) quarks [13]. The precise measurement of this coupling is an important test of the Standard Model prediction. The context of this thesis is the study of $H \to \tau\tau$ decays. The Higgs boson couples in principle to all massive fermions, the coupling strength, however, depends on the mass of the fermion. Since tau leptons are the heaviest leptons, the decay channel $H \to \tau\tau$ is an experimentally promising channel to measure Higgs boson to lepton couplings.

The coupling of τ leptons has been studied in proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 7\,\text{TeV}$ and $\sqrt{s} = 8\,\text{TeV}$ during Run 1 [14, 15]. Two analysis strategies were pursued in parallel. The main strategy used multivariate techniques to increase the sensitivity of the analysis. As a cross check a cut-based analysis was performed.

This thesis focuses on the decay $H \rightarrow \tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ where one τ lepton decays leptonically and the other τ decays hadronically. The thesis uses data from proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 13\,\mathrm{TeV}$ taken with the ATLAS detector during Run 2.

CHAPTER 2

Theory

This chapter introduces the basic theoretical concepts relevant for this thesis. The theoretical framework of the Standard Model of particle physics is Quantum Field Theory. The full rigor of Quantum Field Theory is not necessary to understand the relevant concepts of this thesis, therefore an attempt has been made to keep the mathematical formalism as simple as possible. The material in this chapter is based on Refs. [10] and [16], the presentation of the material follows Ref. [10].

The first section gives an overview of the Standard Model of particle physics. It outlines its structure and its concepts. The following section focuses on Quantum Electrodynamics (QED) and illustrates the importance of symmetries in particle physics. Section 2.3 describes how to include weak interactions in the same model. Section 2.4 introduces spontaneous symmetry breaking and shows how this mechanism can be exploited to give mass to the gauge bosons. The chapter closes with Sec. 2.5, which applies the principle of spontaneous symmetry breaking also to generate mass terms for charged fermions.

2.1 The Standard Model

The Standard Model of particle physics (abbreviated as SM in this thesis) was developed during the second half of the 20th century. The SM describes three of the four known forces, namely the strong force, electromagnetic force and the weak force. The Standard Model predicts the evolution and interaction of all known particles governed by these three forces. The effect of the fourth force, gravitation, is currently not measurable at the microcosm of elementary particles. Gravitation is not incorporated in the Standard Model of particle physics.

The elementary particles of the Standard Model can be arranged into two groups: *bosons*, with integer spin, and *fermions*, with half-integer spin. The group of fermions consists of two types of particles: *leptons* and *quarks*. The matter that surrounds us is made of protons (p), neutrons (n) and electrons (e). Protons and neutrons are composite particles and consist of up (u) and down (d) quarks. The electron is an elementary particle and belongs to the group of leptons. The electron, the up and

Table 2.1: List of all particles in the Standard Model. Masses are in MeV, charge in multiples of the proton charge. The generations are separated by horizontal lines for quarks and leptons. Values are rounded and are taken from Ref. [17], except the Higgs boson mass $m_H = 125.09 \pm 0.21(\text{stat.}) \pm 0.11(\text{syst.})\,\text{GeV}$, which is taken from Ref. [13]. The number of printed digits does not correspond to uncertainty. Only an upper limit is shown for the neutrino masses.

	Particle	Charge	Mass
Quarks	up u	$+2/3$	2.2
	down d	$-1/3$	4.8
	charm c	$+2/3$	$1,275$
	strange s	$-1/3$	95
	top t	$+2/3$	$173,210$
	bottom b	$-1/3$	$4,180$
Leptons	electron e^-	-1	0.511
	e neutrino ν_e	0	$< 2 \cdot 10^{-6}$
	muon μ^-	-1	106
	μ neutrino ν_μ	0	< 0.19
	tau τ^-	-1	$1,777$
	τ neutrino ν_τ	0	< 18.2
Bosons	gluon g	0	0
	photon γ	0	0
	$W^\pm$ boson	± 1	$80,385$
	Z boson	0	$91,188$
	Higgs boson H	0	$125,090$

the down quark together with the electron neutrino comprise the *first generation* of particles. Each particle in the first generation has a heavier sibling particle in the second and third generation with the same properties, but different mass. The heavier siblings of the electron are the muon (μ) and the tau (τ). The muon belongs to the second generation, the tau to the third generation. The particles of the SM and their properties are listed in Tab. 2.1. Forces between fermions are mediated by the exchange of bosons: gluons (g) for the strong force, photons (γ) for the electromagnetic force, and W and Z bosons for the weak force. Ultimately, there is the long-sought-for Higgs boson (H), which is a necessary ingredient of the SM. Fermions, heavy gauge bosons and the Higgs boson itself acquire their mass by interacting with the Higgs field.

The theory of QED is a part of the SM. QED describes the electromagnetic interactions of all charged particles. The force is mediated by the exchange of photons. QED is introduced in Sec. 2.2 and extended in Sec. 2.3 to incorporate also weak interactions mediated by the Z and W bosons. Besides the electroweak sector of the SM, there is Quantum Chromodynamics (QCD) which describes strong interactions mediated by gluons. In QCD a color charge is introduced for quarks. That means quarks come in three different colors: red, green and blue. Since QCD is not crucial for the BEH mechanism for leptons, QCD is not discussed in this thesis.

Symmetries play a fundamental role in the structure of SM and guide the construction of interactions within the model. All three forces can be associated with a symmetry group. The interactions are invariant under the local transformations of the symmetry groups. The full symmetry group of the Standard Model is

$$\underbrace{SU(3)_C}_{\text{QCD}} \times \underbrace{SU(2)_Y \times U(1)_L}_{\text{electroweak}} \tag{2.1}$$

where $SU(n)$ denotes the special unitary group of dimension n and the $U(1)$ denotes the one-dimensional unitary group. The indices will become apparent in the course of this chapter. The symmetry group of electroweak interactions is detailed in Sec. 2.3.

The success of the SM continued in 2012 with the discovery of the Higgs boson by the ATLAS and CMS collaborations at CERN (European Organization for Nuclear Research) [3, 4]. The predictions of the SM have been verified in various analyses and experiments with remarkable precision. Despite its success, the SM does not explain all of the observed phenomena. For example, it does not predict important parameters of the theory, such as the masses of the particles. Furthermore the observation of neutrino oscillation indicates that neutrinos have non-zero mass. Massive neutrinos, however, are forbidden in the SM. There are other models or extensions to the SM, to solve some of the problems, but these models are beyond the scope of this thesis.

2.2 Quantum Electrodynamics

Quantum Electrodynamics describes the interaction of electrically charged particles with photons. The theory is formulated as a Quantum Field Theory (QFT). As the name indicates a central part of QFT are fields. Excitations of fields correspond to particles. The dynamics can be derived from the Lagrangian density $\mathcal{L}$. Strictly speaking the term Lagrangian refers to $L = \int \mathrm{d}x^3 \mathcal{L}$, but it is customary to use Lagrangian also for the Lagrangian density. The Lagrangian density takes the role of the Lagrangian in classical mechanics. Since the topic of this thesis is about $H \to \tau\tau$ decays, the Lagrangian

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi. \tag{2.2}$$

of a free, *i.e.*, non-interacting, tau lepton of mass m is considered. The field associated with the tau is ψ. The tau lepton is a fermion, therefore the field ψ is a four-component object called *spinor*. A spinor should not be confused with a Lorentz four-vector. The equation of motion for the field ψ can be derived by using the quantum field theoretical analogous of the Euler–Lagrange equation

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial\left(\partial_\mu \psi\right)} \right) - \frac{\partial \mathcal{L}}{\partial \psi} = 0. \tag{2.3}$$

Without loss of generality ψ and its adjoint spinor $\bar{\psi} = \psi^{\dagger}\gamma^{0}$ can be treated as independent variables, and thus Eq. (2.3) can be used for $\bar{\psi}$ to derive the equation of motion of the free tau lepton

$$(i\gamma^{\mu}\,\partial_{\mu} - m)\psi = 0 \tag{2.4}$$

also know as the Dirac equation.

The previous example considered a non-interacting tau lepton. QED describes how charged particles interact via the exchange of photons. To incorporate the photon field A, we can use a handy tool: *local gauge symmetries.* Symmetries have always been a useful tool in physics, especially in high energy physics. Maxwell's equations of electrodynamics exhibit a global gauge symmetry. That means one is free to choose a global gauge by adding $\partial_{\mu}\Lambda$ to the fields of electrodynamics. In the quantum field theoretical equivalent of Maxwell's electrodynamics, QED, a global gauge symmetry means that the Lagrangian is invariant under global phase transformations $\psi \to \psi' = e^{-iq\lambda}\psi$ of the field, where q is the charge operator. An additional phase in Eq. (2.2) cancels in both $\bar{\psi}\psi$ terms and thus leaves the Lagrangian invariant.

A local gauge symmetry on the other hand means that the phase depends on the position x_{μ}, $\lambda = \lambda(x_{\mu})$ (hence *local*). The Lagrangian presented in Eq. (2.2) is not invariant under the local gauge transformation

$$\psi \to \psi' = e^{-iq\lambda(x)}\psi, \tag{2.5}$$

because the derivation in Eq. (2.2) produces an additional term of $-iq\bar{\psi}\gamma^{\mu}(\,\partial_{\mu}\lambda)\psi$, which does not cancel with any other term in the Lagrangian.

The invariance can be restored with the help of a new field: the vector field A of the photon. The derivative in the Lagrangian is replaced by the covariant derivative $\mathcal{D}_{\mu} = \partial_{\mu} + iqA_{\mu}$. The new field is required to transform according to

$$A_{\mu} \to A'_{\mu} = A_{\mu} + \,\partial_{\mu}\lambda \tag{2.6}$$

under the local gauge transformation. Substituting this into the full Lagrangian yields

$$\mathcal{L} = \underbrace{\bar{\psi}(i\gamma^{\mu}\,\partial_{\mu} - m)\psi}_{\text{free }\tau} - \underbrace{\frac{1}{4\pi}F^{\mu\nu}F_{\mu\nu}}_{\text{free }\gamma} - \underbrace{(q\bar{\psi}\gamma^{\mu}\psi)A_{\mu}}_{\text{interaction}} \tag{2.7}$$

where the term $\frac{-1}{4\pi}F^{\mu\nu}F_{\mu\nu}$ with $F^{\mu\nu} = \partial^{\mu}A^{\nu} - \partial^{\nu}A^{\mu}$ has been added. This term corresponds to the free Lagrangian of the photon field, similarly to equation (2.2) for the tau lepton. This Lagrangian accounts for the dynamics of both fields, the mass m of the tau lepton and the interaction between the tau lepton and the photon. It is instructive to define the current density of QED as

$$j^{\text{em}}_{\mu} = q\bar{\psi}\gamma_{\mu}\psi. \tag{2.8}$$

The interaction term in the Lagrangian corresponds to the contraction of the current density and this photon field.

This final Lagrangian is now invariant under local gauge transformations. By imposing the condition of local gauge invariance, one ends up with the Lagrangian of the tau lepton and the photon field and their interactions. Local gauge symmetries give a guide line on how to introduce the interactions based on a symmetry group. The symmetry transformations in QED belong to the unitary group $U(1)$, *i.e.*, $e^{-iq\lambda} \in U(1)$, where the charge operator q is the generator of the group. This symmetry group is rather simple. The next section extends this idea to a larger symmetry group.

2.3 Electroweak Unification

The Glashow–Weinberg–Salam model (GWS) unifies electromagnetic and weak interactions into a single theory. The previous chapter was based on a local gauge transformation from the 1-dimensional unitary group $U(1)$. For the electroweak unification this symmetry group is extended to

$$SU(2)_L \times U(1)_Y \tag{2.9}$$

where the $SU(2)$ is the 2-dimensional special unitary group. The fermions are cast into left-handed doublets and right-handed singlets. Here left- and right-handedness actually refers to chirality $\frac{1}{2}(1 \pm \gamma^5)$ eigenstates and not to helicity eigenstates. The right-handed singlets[1]

$$e_R, \mu_R, \tau_R, u_R, d_R, \ldots \tag{2.10}$$

do not transform under $SU(2)_L$ symmetry transformations. The left-handed doublets

$$\chi_L = \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L, \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L, \begin{pmatrix} u \\ d \end{pmatrix}_L, \ldots \tag{2.11}$$

are affected by $SU(2)_L$ transformations, hence the subscript L for left-handed.

The other symmetry group of the GWS model, $U(1)_Y$, is similar to the one discussed in the previous section. The difference is that the generator of the symmetry is the hypercharge Y and not the electromagnetic charge q as it was in QED. The hypercharge of the particles in the SM is listed among other properties in Tab. 2.2.

Similarly to the current density j_μ^{em} in Quantum Electrodynamics the GWS model introduces charged weak currents from lepton-neutrino interactions as

$$j_\mu^\pm = \bar{\chi}_L \gamma_\mu \sigma^\pm \chi_L \tag{2.12}$$

using the usual linear combination $\sigma^\pm = \frac{1}{2}(\sigma^1 \pm i\sigma^2)$ of Pauli matrices denoted by σ^i. Motivated by the Gell-Mann–Nishijima formula $q = I^3 + \frac{1}{2}Y$, which relates charge q, third component of isospin I^3 and weak hypercharge Y, one can define a current density of weak hypercharge as

$$j_\mu^Y = 2j_\mu^{\text{em}} - \bar{\chi}_L \gamma_\mu \sigma^3 \chi_L. \tag{2.13}$$

[1] Fermion fields are denoted with f instead of ψ_f in order to avoid cluttering up the notation with indices.

Table 2.2: List of the charge q, the third component of the weak isospin I^3 and the weak hypercharge Y for quarks and leptons. Quantum numbers for the first generation are taken from Ref. [10].

Particles	q	I^3	Y
u_L, c_L, t_L	2/3	1/2	1/3
u_R, c_R, t_R	2/3	0	4/3
d_L, s_L, b_L	−1/3	−1/2	1/3
d_R, s_R, b_R	−1/3	0	−2/3
e^-_L, μ^-_L, τ^-_L	−1	−1/2	−1
e^-_R, μ^-_R, τ^-_R	−1	0	−2
ν_e, ν_μ, ν_τ	0	1/2	−1

The GWS model postulates a coupling of the three weak isospin currents $\boldsymbol{j} = \frac{1}{2}\bar{\chi}_L\gamma\boldsymbol{\sigma}\chi_L$ with $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ to the vector boson field $\boldsymbol{W}$, and a coupling of the weak hypercharge current j^Y to a vector field B, with coupling strength g_w and $\frac{1}{2}g'$ respectively. The relevant term in the Lagrangian reads

$$\mathcal{L}_1 = -i\left[g_w\boldsymbol{j}_\mu \cdot \boldsymbol{W}^\mu + \frac{1}{2}g' j^Y_\mu B^\mu\right]. \tag{2.14}$$

The physical fields of the W boson are $W^\pm = \frac{1}{\sqrt{2}}(W^1_\mu \mp iW^2_\mu)$ and its coupling to the fermion fields can be extracted directly from the Lagrangian. In this theory, the two neutral fields W^3 and B mix to produce the physically observed fields, the photon A and the Z^0

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_w & \sin\theta_w \\ -\sin\theta_w & \cos\theta_w \end{pmatrix} \begin{pmatrix} B_\mu \\ W^3_\mu \end{pmatrix}. \tag{2.15}$$

The angle θ_w is referred to as the *weak mixing angle*. This model is able to explain the observed electromagnetic and weak interactions, including β-decays. The theory had its glorious moment, when the theoretically predicted W and Z bosons were discovered at CERN in 1983 [11, 12].

Despite the success of this model, there is a flaw in the theory. The Lagrangian can not contain a mass term for the Z or W bosons. A term like $\frac{1}{2}m^2B^\mu B_\mu$ breaks gauge invariance under the transformations in Eq. (2.9). The gauge bosons of the GWS model can not have a mass. This is fine for the photon, as it is in fact massless. For the Z and W bosons, however, this is not consistent with experimental measurements. The Z boson mass $m_Z = (91.1876 \pm 0.0021)$ GeV [17] is experimentally measured to high precession and not compatible with a theory of a massless Z boson. The W boson mass is $m_W = (80.385 \pm 0.015)$ GeV [17].

Similarly the Lagrangian can also not contain a mass term for fermions. By using the properties of the chirality operators $\frac{1}{2}(1 \pm \gamma^5)$, the mass term of a tau can be written as

$$-m\bar{\tau}\tau = -m(\bar{\tau}_R\tau_L + \bar{\tau}_L\tau_R). \tag{2.16}$$

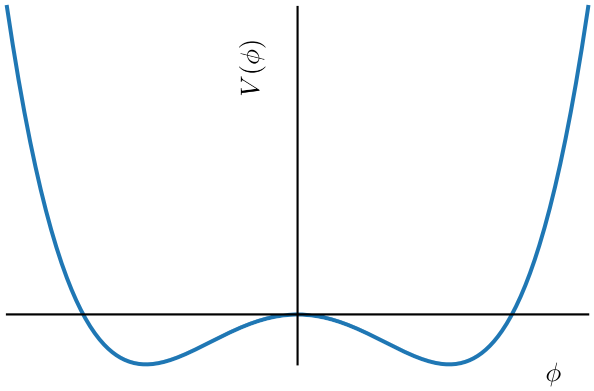

Figure 2.1: Illustration of the potential $V(\phi)$ with $\mu^2 < 0$ and $\lambda > 0$.

In contrast to the right-handed spinors τ_R, which are not affected by the $SU(2)_L$ transformations, the left-handed spinors τ_L are transformed by elements of the $SU(2)_L$ group, which makes this mass term not gauge invariant.

To maintain the gauge invariance of the Lagrangian, another method is required to describe how the gauge bosons and fermions can acquire mass. The solution to this problem is the Brout–Englert–Higgs mechanism by spontaneously breaking the symmetry of the Lagrangian.

2.4 Spontaneous Symmetry Breaking

The idea behind spontaneous symmetry breaking is, that the Lagrangian exhibits a certain symmetry, but the system is not in its ground state in the absence of all field excitations, *i.e.*, $\phi = 0$. The framework of QFT is a perturbative approach, which is valid for small deviations from the ground state of a system. This means in order to use the usual framework, the Lagrangian has to be expanded around its ground state. Using perturbation theory around $\phi = 0$, *i.e.*, not around the ground state of the system, is valid only if one could include infinite orders of perturbation theory. When translated fields are introduced, such that the system is in its ground state in the absence of all field excitations, the original symmetry is still inherent in the Lagrangian, but it is *hidden*. The BEH mechanism, which was suggested by Higgs [6, 7], Brout and Englert [8], Guralnik, Hagen and Kibble [9] exploits this to generate the mass terms for bosons and fermions.

Consider a field ϕ and the potential

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda(\phi^\dagger \phi)^2 \tag{2.17}$$

with $\mu^2 < 0$ and $\lambda > 0$. The potential is sketched in Fig. 2.1. The system is not in its ground state for $\phi = 0$, but is invariant under reflections $\phi \to -\phi$. In order to utilize the field to generate the mass terms for the bosons, ϕ becomes a composite of four scalar real fields ϕ_i

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} . \tag{2.18}$$

The potential $V(\phi)$ does not have a unique minimum, there is rather a manifold, which minimizes the potential. This means there is some freedom in choosing the ground state around which one would like to expand the Lagrangian. The presentation of this topic is especially convenient, if one chooses the ground state as

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} . \tag{2.19}$$

To exploit the spontaneous symmetry breaking to generate the masses of the gauge bosons Z and W, one has to couple the corresponding fields to the field ϕ in the following way:

$$\mathcal{L}_2 = \left| \left(-ig\frac{\boldsymbol{\sigma}}{2} \cdot \boldsymbol{W}_\mu - i\frac{g'}{2} B_\mu \right) \phi \right|^2 \tag{2.20}$$

where $|x|^2 = x^\dagger x$. If this is written out explicitly, the mass term of the W boson becomes apparent with $m_W = \frac{1}{2} vg$. Equation (2.15) can be used on the remainder of equation (2.20). The mass of the Z boson becomes $m_Z = \frac{1}{2} v \sqrt{g^2 + g'^2}$.

By expanding the Lagrangian around its ground state, which is not at $\phi = 0$ due to the functional form of $V(\phi)$, the gauge bosons can acquire mass. This method of spontaneous symmetry breaking to generate the mass terms, is called the Brout–Englert–Higgs mechanism. The last step is to employ the same mechanism to generate mass terms of fermions, for example τ leptons.

2.5 Fermion Masses

In the Standard Model, the same doublet ϕ, which generates the masses of the heavy gauge bosons, can be used to generate the masses of leptons. To achieve this a Yukawa interaction between the lepton fields and the Higgs field is introduced. The relevant term in the Lagrangian for τ leptons is

$$\mathcal{L}_3 = -Y_\tau \left[(\bar{\nu}_\tau, \bar{\tau})_L \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \tau_R + \bar{\tau}_R (\phi^-, \bar{\phi}^0) \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L \right] . \tag{2.21}$$

The combination of the left- and right-handed doublets and the doublet ϕ is invariant under $SU(2)_L \times U(1)_Y$ symmetry transformations. Expanding the field ϕ around the ground state

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \tag{2.22}$$

transforms Eq. (2.21) into

$$\mathcal{L}_3' = -\frac{Y_\tau}{\sqrt{2}} v (\bar{\tau}_L \tau_R + \bar{\tau}_R \tau_L) - \frac{Y_\tau}{\sqrt{2}} (\bar{\tau}_L \tau_R + \bar{\tau}_R \tau_L) h. \tag{2.23}$$

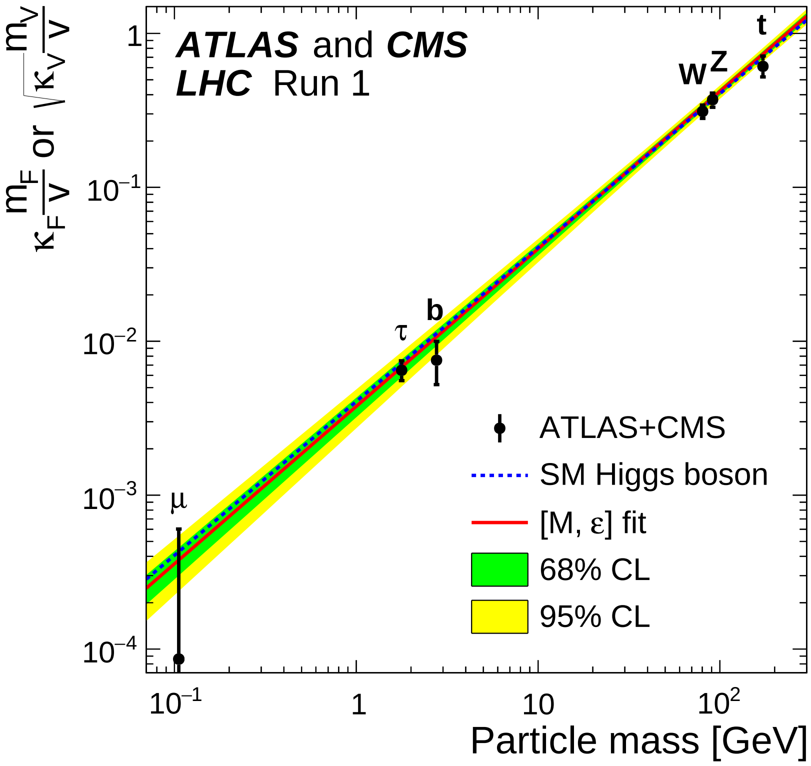

Figure 2.2: Comparison of the measured coupling strengths of different particles to the Higgs boson and the prediction of the Standard Model. Data points use data from the ATLAS and CMS experiments from Run 1. The plot is extracted from [13].

The first term corresponds to the mass term of the lepton with the mass $m = \frac{Y_\tau v}{\sqrt{2}}$. The coupling parameter Y_τ can be chosen such that the predicted mass matches the observed mass. The standard model is thus not able to predict the masses of fermions, they are merely free parameters in the theory, which need experimental input to determine their value.

The second term in $\mathcal{L}_3'$ describes the interaction of the lepton and the Higgs boson. This is the term in the Lagrangian that is responsible for the decay of $H \to \tau\tau$. The coupling strength of the Higgs boson to fermions is therefore proportional to the masses of the fermions. The measurements of the coupling strengths of the Higgs boson to other particles show impressive agreement with the SM prediction. The coupling strength measurements are summarized in Fig. 2.2.

CHAPTER 3

Particle Physics Phenomenology

Many experiments in high energy physics are conducted by colliding particles at high center-of-mass energies. The Large Hadron Collider near Geneva with the ATLAS detector is no exception to this. The analysis methods of the experimental data rely on the comparison of event rates in one way or another. It is therefore instructive to look at some of the phenomena at particle colliders, specifically proton-proton colliders such as the LHC.

The first section of this chapter introduces the concept of cross sections at hadron colliders. Section 3.2 defines the term luminosity and relates it to the event rate of a process. Sections 3.3 and 3.4 describe phenomenological effects encountered at proton-proton scattering at the LHC. The chapter closes with Sections 3.5 and 3.6 with the discussion of the production and decay of the Higgs boson at the LHC.

3.1 Cross Sections and Parton Distribution Functions

Reactions in high energy physics are characterized by their cross section σ. Consider the particles a and b in the initial state and a single final state particle c. Given the Lagrangian of the Standard Model one can derive the cross section $\sigma(a + b \to c)$ for this elementary process. This is detailed in Refs. [10] and [16]. One complication arises in proton-proton collisions by the fact, that the colliding particles are not point-like, elementary particles. In this context the cross section $\sigma(a + b \to c)$ is called *partonic cross section*, because it describes a process at the level of partons, the constituents of the proton. The proton consist of quarks and gluons, therefore the internal structure of the proton has to be taken into account. The initial protons are denoted by A and B and the remnant of the two protons by X. The process at the *hadronic* level reads $A + B \to c + X$. The internal structure of the proton A is quantified by the *parton distribution function* $f_{a,A}(x, Q^2)$, where x is the Bjorken variable, which denotes the fraction of total momentum of the proton that is carried by the parton a, and Q^2 is the squared four-vector of the momentum transferred in this process. The parton distribution function $f_{a,A}(x, Q^2)$ gives the probability to find particle a in A with the momentum fraction x when probing with an interaction of Q^2. Integration over the momentum fractions x_a, x_b of particles a, b and summing over all particles a, b that contribute to $a + b \to c$, yields the *hadronic cross section* [18]

$$\sigma(A + B \to c + X) = \sum_{a,b} \int_0^1 \int_0^1 \mathrm{d}x_a \, \mathrm{d}x_b \; f_{a,A}(x_a, Q^2) \; f_{b,B}(x_a, Q^2) \; \sigma(a + b \to c). \quad (3.1)$$

This approach relies on the concept of factorization and its factorization scale μ_F [19], which determines the separation of the hard scattering processes captured in $\sigma(a+b \to c)$ and the structure of the proton f. The predicted cross sections depend on the factorization scale μ_F and the renormalization scale μ_R, which determines the evaluation of the coupling strength. Both scales are non-physical quantiles and give rise to systematic uncertainties of the predicted cross sections.

3.2 Luminosity

The event rate with which a certain physical process is observed in collider experiments depends on the cross section σ of the process (on the hadronic cross section in proton-proton collision) and on properties of the particle collider. The luminosity $\mathcal{L}$ combines several properties of the collider and the particle beams into a single quantity. Consider two particle beams in a circular collider such as the LHC with a perfectly Gaussian profile of width σ_x and σ_y in the transversal plane. The beams consist of $n_{1,2}$ particles and circulate at a frequency f. The (instantaneous) luminosity L is given by [16]

$$L = f \frac{n_1 n_2}{4\pi\sigma_x\sigma_y}. \quad (3.2)$$

The expected event rate $\dot{N}$ can then be calculated with

$$\dot{N} = \sigma L. \quad (3.3)$$

Integrating over a period of time T and introducing the *integrated luminosity* $L_{\text{int}} = \int_T L \, \mathrm{d}t$, yields the total number of expected events during the period T

$$N = \sigma L_{\text{int}}. \quad (3.4)$$

3.3 Hadronization

The final state of the hard scattering process might contain particles with non-vanishing color charge. Due to the confinement of QCD, these particles can not be observed separately. The colored final state particles have to form color-neutral particles. This process is referred to as hadronization. Perturbation theory breaks down in QCD for low transfered momenta, therefore hadronization can not be calculated within the realm of perturbative QFT. Different methods are used in Monte Carlo generators to simulate this process. One method is to consider the potential energy when separating two particles with color charge. When the energy reaches the threshold of quark-anti-quark creation, new particles are created. This process is repeated until stable particles without color charge are formed [15].

3.4 Underlying Event And Pile-Up

Section 3.1 considered the process of two protons A and B colliding to produce particle c. The final state contains particle c but also the remnants of the two protons X. Particle c is produced in a hard scattering process, whereas the remnants undergo soft scattering. Usually c and X are detected in the experiment. The detector signal of X is referred to as the *underlying event*. As perturbation theory can not be applied in this case, a phenomenological model is employed in the Monte Carlo generators to simulate the underlying event.

At the LHC, the proton beams are partitioned into 2808 bunches, each bunch consists of the order of 10^{11} protons [5]. During each bunch crossing inside the detector, multiple proton-proton collisions take place, which are referred to as in-time pile-up events. With the high instantaneous luminosity during 2016 data taking, the number of interactions per bunch crossing in the ATLAS detector reached values exceeding 35. To model this large number of pile-up interactions, the simulated detector response of inclusive proton-proton events is combined with the actual detector response of the process under study. Details of the simulation process are described in Sec. 4.4.

3.5 Higgs Boson Production

One of the goals of the ATLAS experiment was to discover the Higgs boson. This goal has been achieved in 2012 [3]. At the LHC there are different production mechanisms of the Higgs boson. The production cross sections at the LHC depends on the center-of-mass energy $\sqrt{s}$. The production cross sections can be found in Fig. 3.1. The two most important processes for this analysis are gluon fusion (ggF) and vector boson fusion (VBF). The corresponding leading order Feynman diagrams are shown in Fig. 3.2. The two production mechanisms can be identified by their detector signature.

According to Fig. 3.1 gluon fusion is the dominant Higgs boson production channel at the LHC with a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$. Gluons are present in proton-proton collisions since the strong force, which holds the proton together, is mediated by gluons. In ggF, two gluons couple to a virtual fermion loop. The fermions in the loop annihilate to produce a Higgs boson. The largest contribution comes from virtual top quarks in the loop, since the coupling strength of fermions to the Higgs boson depends on the mass of the fermion and the top quark is the heaviest fermion, see Tab. 2.1. In vector boson fusion, two quarks in the protons radiate virtual $W^\pm$ or Z bosons which fuse to form a Higgs boson. The cross section for VBF at $\sqrt{s} = 13\,\text{TeV}$ is about one order of magnitude lower than that for ggF. However, the process has a crucial role for this analysis, because the VBF production mode has a clear detector signature.

Other Higgs boson production mechanisms at the LHC are VH and $t\bar{t}H$, where $V = W^\pm, Z$. These processes have negligible contributions for this analysis and are therefore not considered.

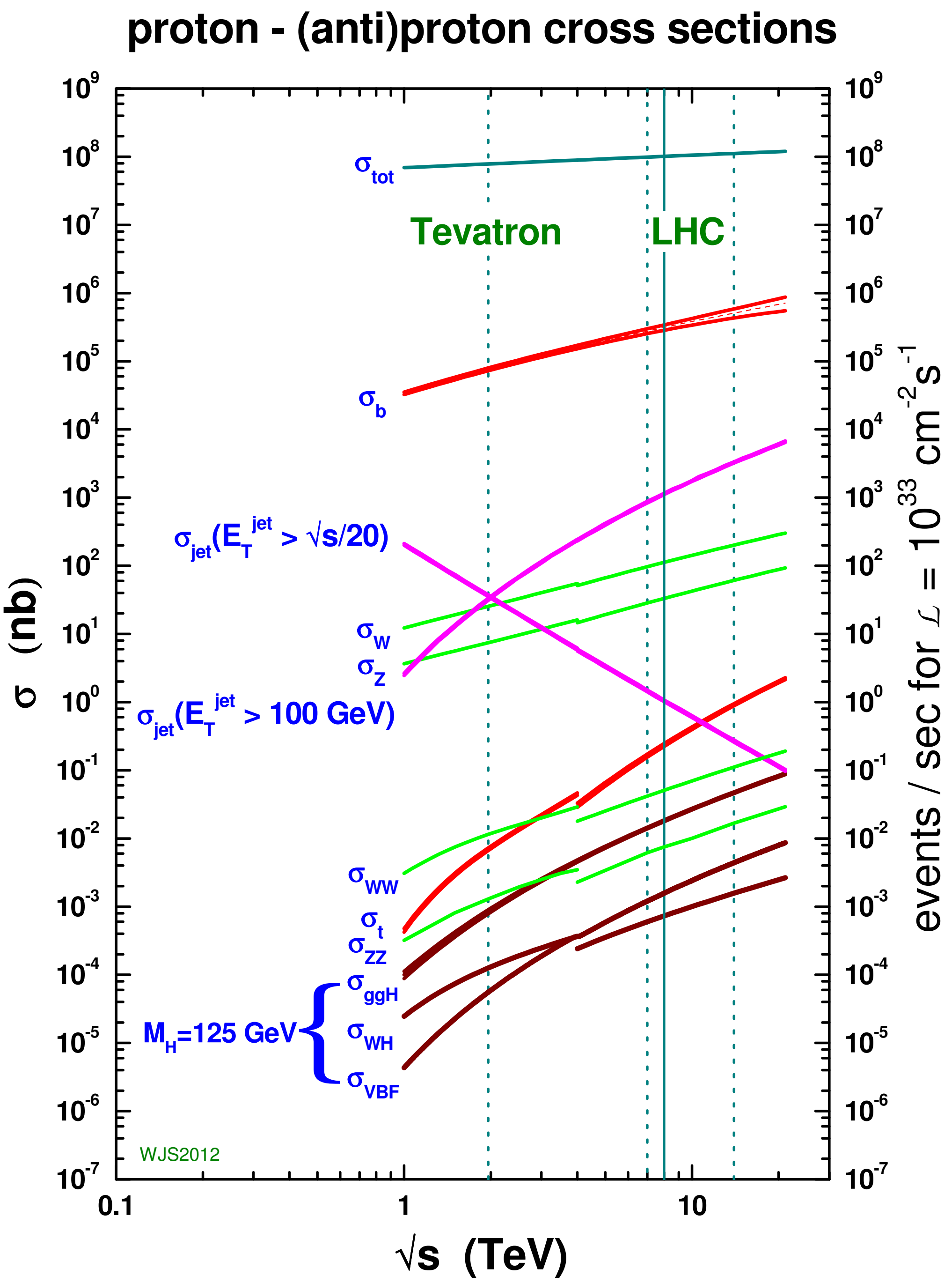

Figure 3.1: Expected cross sections of various processes in proton-(anti)proton collisions as a function of the center-of-mass energy. Below 4 TeV proton-antiproton collisions are assumed, above this threshold proton-proton collisions are shown. The vertical lines indicate the center-of-mass energies $\sqrt{s} = 1.96$ TeV for Tevatron, and $\sqrt{s} = 7$ TeV, $\sqrt{s} = 8$ TeV and $\sqrt{s} = 14$ TeV for the LHC. Gluon fusion is denoted by ggH. Figure taken from Ref. [20].

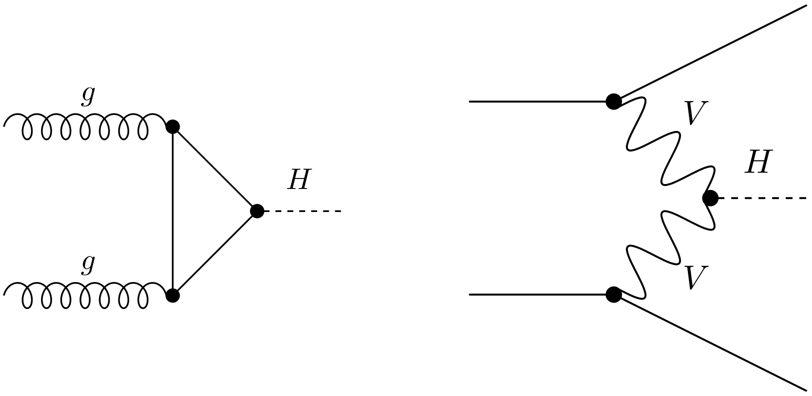

Figure 3.2: Leading order Feynman diagrams of the two most important Higgs boson production mechanisms for this analysis, gluon fusion (left) and vector boson fusion (right).

Table 3.1: Comparison of branching ratios of different decay channels of the Higgs boson with $m_H = 125\,\mathrm{GeV}$ [21].

Decay Channel	Branching Ratio
$H \to b\bar{b}$	$5.824 \cdot 10^{-1}$
$H \to W^+W^-$	$2.137 \cdot 10^{-1}$
$H \to \tau^+\tau^-$	$6.272 \cdot 10^{-2}$
$H \to c\bar{c}$	$2.891 \cdot 10^{-2}$
$H \to ZZ$	$2.619 \cdot 10^{-2}$
$H \to gg$	$2.187 \cdot 10^{-2}$
$H \to \gamma\gamma$	$2.270 \cdot 10^{-3}$

3.6 Higgs Boson Decay

The Higgs boson is an extremely short-lived particle with a total width predicted to be $\Gamma_H = 4.07\,\mathrm{MeV}$ [21]. According to the Standard Model, it can decay into different particles depending on their masses. Important Higgs boson decay channels and their branching ratios are listed in Tab. 3.1.

This thesis focuses on the decay channel $H \to \tau\tau$. Tau leptons couple directly to the Higgs particle. Since they are the heaviest lepton, they present a way to assess the SM coupling of the Higgs boson to leptons. Tau leptons are also short-lived particles with a life time of $(290.3 \pm 0.5) \cdot 10^{-15}\,\mathrm{s}$ [17]. Most tau leptons decay before they reach the detector. The decay of a tau lepton can be grouped into two classes: leptonic decays, where the tau decays into an electron or a muon (and the appropriate neutrinos), and hadronic decays, where the tau decays into quarks (and an (anti)-tau neutrino). The leading order Feynman diagrams of these two processes are shown in Fig. 3.3. Table 3.2 lists the branching ratios of the Higgs boson to the analysis sub-channels lep-lep, had-had and lep-had, where the tau leptons from the decay of the Higgs boson both decay leptonically, both decay hadronically and where one tau decays leptonically and the other one decays hadronically.

Table 3.2: Comparison of branching ratios of different $H \to \tau\tau$ decay sub-channels. Branching ratios calculated from the tau-to-leptons branching ratio of 35.24% [17].

Decay Sub-Channel	Branching Ratio
$H \to \tau_{\text{lep}}\tau_{\text{lep}}$	12.4 %
$H \to \tau_{\text{lep}}\tau_{\text{had}}$	45.6 %
$H \to \tau_{\text{had}}\tau_{\text{had}}$	41.9 %

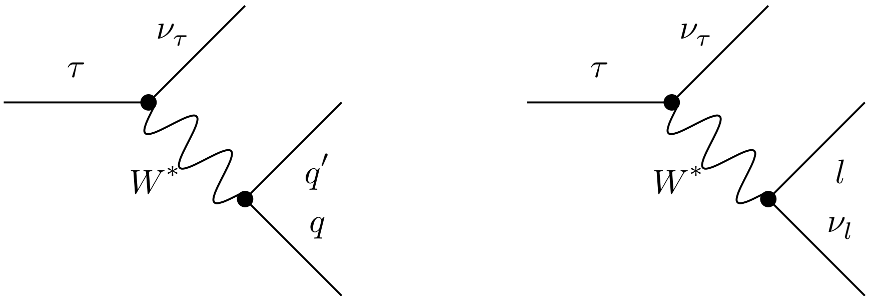

Figure 3.3: Feynman diagrams of hadronic (left) and leptonic (right) tau lepton decay. Particles and anti-particles are not distinguished in this figure, the diagrams exist for τ^+ and τ^- decays.

CHAPTER 4

The Large Hadron Collider and the ATLAS Experiment

Over the last century the field of high energy physics was born and the experiments and the necessary experimental effort grew. The research center CERN in Geneva hosts several large scale experiments, one of which is the Large Hadron Collider, which is able to deliver proton-proton collision at unprecedented center-of-mass energy. One of the large-scale experiments at CERN is ATLAS. The data used in this thesis has been recorded by the ATLAS detector. Simulated events used in this thesis have been processed within the ATLAS collaboration. This chapter gives a brief overview over the Large Hadron Collider in Sec. 4.1 and the ATLAS experiment in Sec. 4.2. Section 4.3 outlines how particles are identified and reconstructed. The chapter closes with a brief discussion of how Monte Carlo events are produced in Sec. 4.4.

4.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is a particle accelerator near Geneva [5]. The collider is capable of accelerating heavy ions, such as lead, and protons. As this thesis studies only proton-proton collision, the collisions involving lead, are not considered. At the time of writing the LHC is the particle accelerator operating at the highest center-of-mass energy. The design energy for proton-proton collisions is $\sqrt{s} = 14\,\text{GeV}$, while the data analysed in this thesis were collected at $\sqrt{s} = 13\,\text{GeV}$, the highest thus far achieved center-of-mass energy of any collider in the world. The accelerator is housed in an underground tunnel, which was formerly used for the Large Electron-Positron (LEP) collider. The circumference of the LHC is about 27 km.

The high center-of-mass energy at the LHC is achieved by using several accelerator stages, which successively increase the energy of the particles. The final stage, the LHC, consists of two evacuated circular beam pipes. Both proton beams are bent by superconducting dipole magnets. The strength of the magnetic fields is the limiting factor for the center-of-mass energy. Each beam is partitioned into 2808 bunches each consisting of approximately 10^{11} protons. The two beams are brought

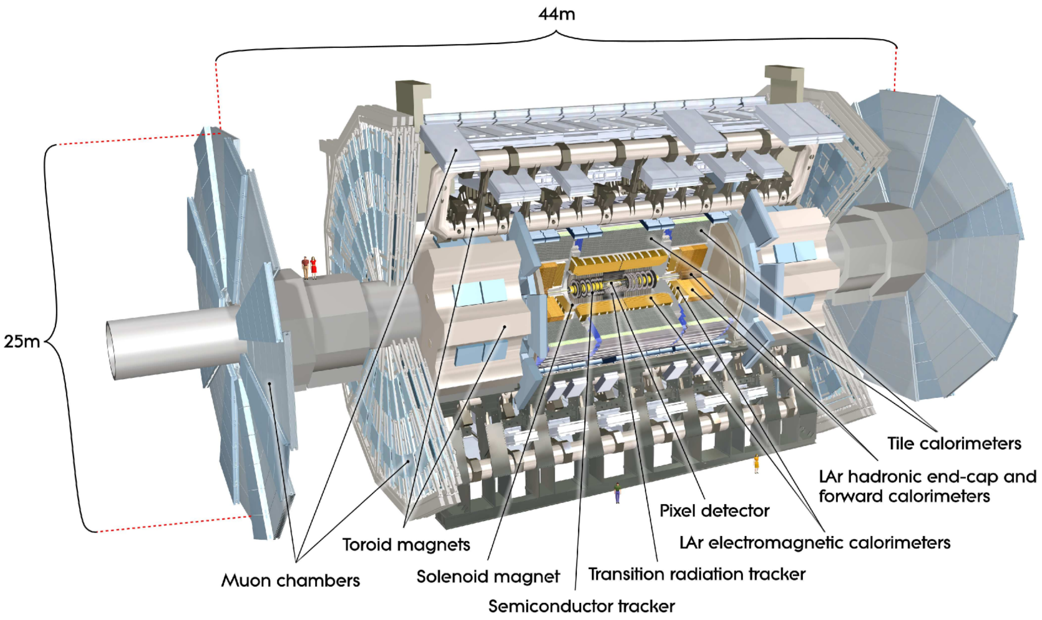
44m
25m
Tile calorimeters
LAr hadronic end-cap and forward calorimeters
Pixel detector
LAr electromagnetic calorimeters
Toroid magnets
Transition radiation tracker
Muon chambers
Solenoid magnet
Semiconductor tracker

Figure 4.1: Sketch of the ATLAS detector and its components [1].

in collision at four interaction points.

Detectors are built around each collision point, to measure the interaction products. The two large, general-purpose experiments at the LHC are ATLAS and CMS. They are designed to study fundamental interactions and to search for new physics. Two other large-scale experiments at the LHC, namely ALICE and LHCb, investigate interactions of heavy ions and decays involving b-hadrons, respectively.

4.2 The ATLAS Experiment

The ATLAS collaboration built, maintains, and operates the ATLAS detector, which is a general-purpose detector at the LHC. The detector nearly covers a solid angle of 4π around the interaction point. The cylindrical detector design has a forward-backward symmetry and the axis of the cylinder is aligned with the beam axises. Its length is 44 m with a diameters 25 m. The detector is split into several different components. The primary detector consists of an inner detector, calorimeters and muon spectrometer. Due to the high instantaneous luminosities at the LHC, a special trigger system has to be used, to select the events, which are permanently recorded. The different parts of the detectors are summarized in the following sections. A more detailed description can be found in [1]. In the course of the detector upgrade between Run 1 and Run 2, some details of the detector have been changed. The description here considers the state of the detector after the upgrade, as it was used to collect the data for this thesis. Figure 4.1 shows an illustration of the ATLAS detector.

4.2.1 Coordinate System

The ATLAS collaboration employs a right-handed Cartesian coordinate system, its z-axis is aligned with the beam axis. The x-axis is pointing to the center of the LHC ring and the y-axis is pointing upwards. It is useful to introduce a coordinate system (r, ϕ, θ) with $r = \sqrt{x^2 + y^2}$, $\phi = \arctan_2(x, y)$ and $\theta = \arctan_2(r, z)$. The interaction point is located at the center of the coordinate system. The pseudorapidity η is introduced as $\eta = \log \tan(\theta/2)$. If the quantities are used to express momenta, the pseudorapidity η equals the rapidity y for massless particles. The rapidity has the nice property, that rapidity differences Δy are invariant under Lorentz boosts along the z-axis, therefore it is customary to use the coordinate system (r, ϕ, η).

4.2.2 Inner Detector

The inner detector (ID) is the part of the detector closest to the beam pipe. The inner detector can be further split into the Pixel Detector, Semiconductor Tracker (SCT) and Transition Radiation Tracker (TRT). The inner detector is emerged in a 2 Tesla magnetic field from a superconducting solenoid around the inner detector. Charged particles are bent due to the magnetic fields. From the hit positions of the particles, it is possible to calculate the curvature of the particle trajectory. The inner detector is therefore a valuable tool to measure the momentum of charged particles.

The silicon Pixel Detector consists of four cylindrical layers and covers the radial extension up to 12 cm. The forward regions up to $|\eta| < 2.5$ are covered by three disks. The first layer of the Pixel Detector, and therefore the layer closest to the interaction point, is the Insertable B-Layer (IBL) [22]. Its radial extension covers the range from about 3 cm to 4 cm around the particle beam. This layer has been installed in 2014 between Run 1 and Run 2 [23]. The IBL provides important input to identify b hadrons, as they travel a measurable distance before they decay, and thus form a secondary vertex. The Pixel Detector consists of about 80 million pixels in total with a minimal size of $50 \times 250\,\mu\text{m}^2$.

The SCT encloses the Pixel Detector. It consists of four layers in the central region and nine disks in the forward regions covered with silicon strip detectors. The accuracy of the SCT in the barrel region is $17 \times 580\,\mu\text{m}^2$.

The Transition Radiation Tracker is made of drift tubes and polypropylene and polyethylene fibers. The TRT is used to measure primary ionization from traversing charged particles and to measure transition radiation, which occurs when charged particles pass materials with different dielectric constants. The TRT covers the pseudorapidity range of $|\eta| < 2$. It provides only information in the ϕ direction with a accuracy of $130\,\mu$m.

The tracking detector is designed [1] to have a momentum uncertainty of

$$\frac{\sigma_{p_\text{T}}}{p_\text{T}} = 1\% \oplus 0.05\% \cdot p_\text{T} \cdot \text{GeV}^{-1}, \tag{4.1}$$

where $a \oplus b$ denotes the addition in quadrature $\sqrt{a^2 + b^2}$.

4.2.3 Calorimeter

The calorimeters are used to measure the energy of particles by absorbing them. The ATLAS detector features two types of calorimeters. The inner one, the electromagnetic calorimeter, is intended to contain showers initiated by electrons and photons. Hadrons deposit only a fraction of their total energy in the electromagnetic calorimeter. The hadronic calorimeter encloses the electromagnetic calorimeter and is intended to fully absorb the shower initiated by hadrons and to measure their energy. The shape of the energy deposition in both calorimeters depends on the particle type and therefore provides information for particle identification.

The electromagnetic and the hadronic calorimeter are sampling calorimeters. The active medium of the electromagnetic calorimeter is liquid argon (LAr), which is interleaved with lead absorbers. The calorimeter covers the barrel region up to $|\eta| < 1.475$. The thickness of the electromagnetic calorimeter is greater than 22 radiation lengths (X_0) in the barrel region. The end cap wheels cover pseudorapidity ranges up to $|\eta| < 3.2$. The electromagnetic calorimeter provides a fine segmentation of up to $(\Delta\eta, \Delta\phi) = (0.0031, 0.1)$ in the inner most layer of the barrel region. This provides important information for the identification of $\pi^0 \to \gamma\gamma$ decays, which occur in the decay of τ leptons. The relative energy resolution of the electromagnetic calorimeter is designed [1] to be

$$\frac{\sigma_E}{E} = 0.7\% \oplus \frac{10\%}{\sqrt{E \cdot \mathrm{GeV}^{-1}}}. \tag{4.2}$$

The hadronic calorimeter is built around the electromagnetic calorimeter. It consists of three parts: the hadronic tile calorimeter in the barrel region, the hadronic end caps and the liquid argon forward calorimeter. The barrel region is a sampling calorimeter of scintillating tiles and steel absorbers and covers the pseudorapidity range $|\eta| < 1.7$. The end cap wheels are sampling calorimeters with liquid argon as active medium and interleaved copper absorbers. The end caps cover the range $1.5 < |\eta| < 3.2$. The liquid argon forward calorimeters cover the pseudorapidity range from 3.1 to 4.9. It is a combination of electromagnetic and hadronic calorimeter. The first layer is a sampling calorimeter with copper absorbers for electromagnetic showers, whereas the second and third layer uses tungsten as absorber material for hadronic showers.

The granularity of the hadronic calorimeter is $(\Delta\phi, \Delta\eta) = (0.1, 0.1)$ for $1.5 < |\eta| < 2.5$ and is therefore coarser compared to the electromagnetic calorimeter. The granularity is expected to be sufficient, since hadronic showers tend to be wider compared to electromagnetic showers. The barrel region extends from an inner radius of 2.28 m to an outer radius of 4.25 m, which corresponds to 9.7 interaction lengths (λ). The design energy uncertainty [1] for hadronic jets of the barrel region and the end cap calorimeters is

$$\frac{\sigma_E}{E} = 3\% \oplus \frac{50\%}{\sqrt{E \cdot \mathrm{GeV}^{-1}}}. \tag{4.3}$$

4.2.4 Muon System

The Muon System (MS) constitutes the outer most part of the detector. A toroidal magnet provides a magnetic field to bend the trajectory of the muons. Based on the curvature a momentum measurement between 3 GeV and a few TeV is possible. The barrel region of the Muon System provides coverage of the pseudorapidity region $|\eta| < 1.7$. Four end cap wheels extend this coverage up to $|\eta| < 2.7$. Several different detector technologies are used. The barrel region consists of monitored drift tubes which have a spacial resolution of 35 μm. In detector regions, where the total particle flux exceeds the limits of the monitored drift tubes, cathode strip chambers are used. Additionally resistive plate chambers are used for triggering purposes. The end cap wheels are also equipped with thin gap chambers. The design momentum resolution [1] is

$$\frac{\sigma_{p_\mathrm{T}}}{p_\mathrm{T}} = 10\% \tag{4.4}$$

for muons with transverse momentum $p_\mathrm{T} = 1$ TeV. In order to achieve this the Muon System must reach position measurement accuracies in the z direction of 50 μm or better.

4.2.5 Trigger

Due to the high instantaneous luminosity and the very large total pp cross section at the LHC, see Fig. 3.1, it is not possible to store the detector response for every single bunch crossing. In order to reduce the event rate from 40 MHz to about 1 kHz two different trigger stages are used: the level 1 (L1) trigger and the high level trigger (HLT).

The first level consists of specialized hardware, which takes information from the Muon System and the calorimeters as inputs. The first level trigger searches for regions of interest in the detector response, which indicate an event relevant for physics analyses. The trigger either discards an event or forwards the information to the high level trigger within 2.5 μs. The detector sub-systems are read out, if the L1 trigger accepts an event. At this stage the event rate is reduced to about 100 kHz.

The second stage, the high level software trigger, is implemented on commercially available hardware. The high level trigger uses the information gathered by the L1 trigger, but has also access to more information from the detector including the tracking detectors. The high level trigger performs a full event reconstruction. The reconstruction process is similar to the offline event reconstruction. The HLT takes the final decision within 0.2 s, whether an event should be stored, and achieves the desired event rate of about 1 kHz. Approximately 1 GB/s of data is written to disk.

4.3 Reconstruction and Particle Identification

Measured data events consist of recorded measurements of the detector response. As outlined in Sec. 4.4, the detector response for Monte Carlo events is simulated. The information from data and Monte Carlo is fed to the reconstruction algorithms in order to identify the physical particles in the event and their properties as measured by the detector. The description of the reconstruction is based on Ref. [15].

The reconstruction starts by examining the hits of the inner detector. Reconstructed hits are subject to a Kalman filter to search for tracks. The identified tracks are subject to track quality criteria. Tracks with too few hits are discarded. Several properties of the remaining tracks are calculated including the origin of the track in the z direction. Vertex candidates are chosen from the distribution of z positions of the tracks. The vertex with the largest sum of squared transverse momenta ($\sum p_{\mathrm{T}}^2$) is the primary vertex.

The energy deposits in the electromagnetic and hadronic calorimeter are subject to clustering algorithms. The algorithm is required to be invariant under collinear and infrared radiation of particles in a jet. The topological clustering is seeded with calorimeter cells, whose signal exceeds the threshold t_{seed}. Neighbouring cells are added to the topological clusters if their signal exceeds the threshold $t_{\mathrm{neighbour}}$. Furthermore all cells whose signal exceeds t_{cell} are added to the most significant, adjacent seed cluster. The jet reconstruction uses TopoClusters built with $(t_{\mathrm{seed}}, t_{\mathrm{neighbour}}, t_{\mathrm{cell}}) = (4, 2, 0) \cdot \sigma$, where σ is the expected noise in the calorimeter cells. An anti-k_t [24] algorithm is used to reconstruct jets based on the TopoClusters. The anti-k_t algorithm introduces the the distance measures

$$d_{ij} = \min\left((p_{\mathrm{T}}^i)^{-2}, (p_{\mathrm{T}}^j)^{-2}\right) \frac{(\Delta R^{ij})^2}{r^2}, \tag{4.5}$$

$$d_i = (p_{\mathrm{T}}^i)^{-2} \tag{4.6}$$

for the clusters i, j, where ΔR^{ij} is the geometrical distance between cluster i and j in the η-ϕ plane, and p_{T}^i is the transverse momentum associated with cluster i. The clustering depends on the free parameter r. The distance measures d_i and d_{ij} are evaluated for all clusters and all pairs of clusters, respectively. If the minimal distance is of type d_i, the algorithm declares cluster i as a jet and removes cluster i from the list of clusters. If the minimal distance is of type d_{ij}, the algorithm merges clusters i and j. The procedure is repeated, until all clusters are merged or declared as jets.

Jets initiated by b hadrons can be identified due to the long lifetime of b hadrons. The decay products of the b hadrons originate from a secondary vertex in a measurable distance from the primary vertex. This is commonly referred to as b-tagging and can be used to suppress backgrounds including b hadron decays.

The electron reconstruction uses a sliding window algorithm to search for energy clusters in the electromagnetic calorimeter with an associated track in the inner detector. Several criteria including the shower shape information, the quality of the track and the quality of the track-to-cluster association are used to discriminate against other background processes such as electrons from photon conversion in the inner detector. Three working points are defined, namely tight, medium and loose, which correspond to a signal efficiency of 75%, 85% and 95%, respectively.

Muons escape the detector with minimal energy deposition in the detector. Electrons and photons are usually contained in the electromagnetic calorimeter. The hadronic calorimeter is intended to contain hadronic showers. Therefore a track in the Muon System is a clear signature of a muon in the event.

The detector signature of a tau depends on its decay mode. The decay of a tau lepton can be grouped into two classes: leptonic decays and hadronic decay. The corresponding Feynman diagrams are shown in Fig. 3.3.

In case of a leptonically decaying tau, the tau decays directly to an electron or muon and two neutrinos. Neutrinos do not interact with the detector and therefore escape the detector unmeasured. Since the neutrinos carry momentum, the signature of these neutrinos is missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$. Missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}} = -\left|\sum_i \boldsymbol{p}_{\mathrm{T}}^i\right|$ is the momentum required to balance the transverse momenta $\boldsymbol{p}_{\mathrm{T}}^i$ of all visible products i in the event. This includes soft tracks and the products of the hard scattering process. The electron or muon can hardly be distinguished from prompt electrons or muons. A leptonic decaying tau presents itself as a light lepton and is therefore often simply referred to as lepton.

In case of a hadronically decaying tau, the tau decays to quarks and a neutrino. The presence of the neutrino also leads to missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$ similar to the leptonic case. The quarks hadronize and in most cases form charged $\pi^{\pm}$ and neutral pions π^0 [17]. The pions are reconstructed as jets in the calorimeters. The shower shape, however, is different if the shower was initiated by a tau or by QCD processes. Showers from taus are more collimated due to the relatively small mass of the tau compared to its large momentum and the low number of resulting hadrons. A boosted decision tree (BDT, see Sec. 6.1.2) is trained to discriminate against jets. Similarly to the electron, three signal efficiency working points tight, medium and loose are defined. The working points correspond to a reconstruction and identification efficiency of 45%, 55% and 60% for 1-prong, and 30%, 40% and 50% for 3-prong decays [25].

4.4 Monte Carlo Generator and Detector Simulation

The production chain for Monte Carlo events can be divided into three steps. This devision into three steps is not universal. The details of the individual steps are beyond the scope of this thesis, only a brief overview of the concepts is described here. The three steps are: event generation, simulation of the detector response and digitization.

The first step is the actual event generation. This step includes calculating the matrix element of the simulated process. Decays of intermediate particles, which decay in-flight, are simulated to resemble the physical processes in the detector. This step also comprises the hadronization mentioned in Sec. 3.3. Additionally, the radiation of additional particles in the initial (ISR) and final (FSR) state is considered. The output of this stage is a tree-like[1] decay structure of particles. The tree indicates the decay products of intermediate particles, similarly to an family tree. All the particles are annotated with their four-vector momentum. The kinematics of the events are therefore completely determined after the event generation.

The next step in the chain of Monte Carlo production is usually to simulate the response of the detector material. For this, a complete and detailed three-dimensional model of the detector is used. This step comprises the simulation of detector material and the interaction of the decay products of the pp scattering with the detector material. The output of this step is a precise map of energy depositions, commonly referred to as *hits*.

There are different algorithms to simulate the interactions of the inclining particles and the detector material. The procedure called *Full Simulation* (FS) performs a very detailed probabilistic simulation, taking all secondary particles into account,

[1] The complexity of this structure is not limited to a tree. I fact, a simple reaction such as $A + B \rightarrow C* \rightarrow D + E$ can not be represented in a tree like structure. The side of the decay products, however, in this case D and E is usually limited to a tree-like structure.

which are produced in interactions of generated or other simulated particles with the detector material. The full simulation of the detector is time consuming and takes of the order of ten minutes for a single event on a single CPU core.

An alternative approach is to parameterize the average detector response as a function of the momentum of the particles, and thus avoid to simulate all the secondary particles independently. This approach is called *ATLFAST-II* (AF-II). A single event can be simulated within about 10 seconds on a typical CPU core.

Next in the production chain is the digitization. As described in Sec. 3.4 several scattering processes occur during one bunch crossing. This is not considered in the event generation step. To account for pile-up, inclusive pp interactions have been generated and simulated. These samples are intended to not introduce any bias in the process composition and are therefore called *minimum bias* events. The hit information from minimum bias events is merged with the hit information of the process under study.

The detector components measure an analog signal (for example by collecting charge) and turn this signal into a digital value, which can then be processed and stored on disk. The digitization step performs the conversion from (pseudo) analog signals to a digital values by simulating the read out systems.

The generated Monte Carlo after the digitization step corresponds to measured data recorded with the detector. The reconstruction and derivation is repeated for Monte Carlo to be able to compare the measured distributions to the Monte Carlo expectation.

CHAPTER 5

Analysis Strategy

Similarly to the analysis strategy in Run 1, the analysis presented in this thesis pursues two different kinds of analysis methods: a cut-based approach and a multivariate approach. The multivariate analysis (abbreviated as MVA) in Run 1 provided a higher sensitivity and was ultimately selected as the main analysis method. The cut-based analysis (abbreviated as CBA) was used a backup and cross check [14].

The cut-based analysis method uses conditions on kinematic variables to select a region in phase space where the signal contribution is enhanced. Usually the cuts are simple threshold or cutoff conditions on kinematic variables, such as p_{T}. By doing this, the selected region is a rectangular volume in phase space. The possibilities of selecting a specific region in that way are therefore geometrically limited. The selection of cuts is usually guided by a combination of physical insight and studies based on Monte Carlo events to optimize the cuts. The advantage of the cut-based analysis is, that the selection of a region in phase space is directly apparent from the cut flow. The cuts are applied sequentially and studied separately as opposed to a multivariate analysis.

The multivariate analysis of this thesis also uses cuts on the kinematic variables to select a region in phase space, similar to a cut-based analysis, but the selected volume is much larger and the signal-over-background ratio does not reach the purity from the cut-based analysis. Multivariate methods are used within this volume. This means that the kinematic variable space is considered as a whole [19], as opposed to the cut-based analysis which considers the cuts sequentially. This analysis method is therefore able to employ correlations between kinematic variables of the event and select much more complex structures in phase space. The multivariate analysis is closely related the field of *machine learning* and *pattern recognition.* An algorithm is trained on Monte Carlo events. During the training the algorithm learns how to identify signal-like regions in phase space. When the algorithm is used on real data the algorithm can check if the event belongs to this interesting region and therefore classify events whether they are signal-like or background-like. The details of the multivariate analysis algorithms are shown in Sec. 6.1.

This thesis presents both analysis methods and compares their results. Both

Table 5.1: Summary of all triggers used for this analysis. Different single lepton triggers are used in 2015 and 2016.

Lepton Flavor	Year	Trigger

Electron	2015	`HLT_e24_lhmedium_L1EM20VH`
		`HLT_e60_lhmedium`
		`HLT_e120_lhloose`

	2016	`HLT_e24_lhtight_nod0_ivarloose`
		`HLT_e60_lhmedium_nod0`
		`HLT_e140_lhloose_nod0`

Muon	2015	`HLT_mu20_iloose_L1MU15` `HLT_mu40`
	2016	`HLT_mu24_imedium` `HLT_mu50`

analysis methods share the same triggers and data taking periods shown in Sec. 5.1, the same analysis specific kinematic variables explained in Sec. 5.2, the same background and signal model detailed in Sec. 5.3, the same preselection cuts listed in Sec. 5.4 and the same control regions as defined in Sec. 5.7. The analysis method specific event categorization is presented in Sec. 5.5 for the cut-based analysis and in Sec. 5.6 for the multivariate analysis.

5.1 Data Taking

The analysis presented in this thesis uses data recorded with the ATLAS detector during Run 2 in 2015 and 2016. The analysis is restricted to datasets where all detector subsystems were operational and the Large Hadron Collider bunch spacing was 25 ns. To maintain data quality, only events with a primary vertex, which has at least two associated tracks, are considered. The analysis uses *single lepton triggers* for electrons and muons. The trigger properties and thresholds differ between 2015 and 2016 due do different instantaneous luminosities. The triggers used for this analysis are summarized in Tab. 5.1. An overall logical OR of the triggers is used to select the events. An event triggered by any of the triggers is used for the analysis. The leading lepton in the analysis is required to match the lepton, which caused the trigger to fire.

The integrated luminosity of data taken in 2015 is $L_{\mathrm{int}}^{2015} = 3.2\,\mathrm{fb}^{-1}$. The contribution from data taken in 2016 is $L_{\mathrm{int}}^{2016} = 10.0\,\mathrm{fb}^{-1}$, which gives a total integrated luminosity of $L_{\mathrm{int}} = 13.2\,\mathrm{fb}^{-1}$.

5.2 Analysis Specific Quantities

The analysis defines several quantities derived from kinematic variables of an event. Some of the derived quantities are exclusively used in the multivariate analysis, others are specific to the cut-based analysis. The remainder of this section summarizes important quantities used in this analysis.

Missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}} = -\left|\sum_i \boldsymbol{p}_{\mathrm{T}}^i\right|$ is defined as the transverse momentum that is needed to balance the visible components of the event in the transverse plane.

Transverse mass m_{T} of the lepton and $E_{\mathrm{T}}^{\mathrm{miss}}$ is defined as

$$m_{\mathrm{T}} = \sqrt{2\, p_{\mathrm{T}}^{\mathrm{lep}}\, E_{\mathrm{T}}^{\mathrm{miss}}\, (1 - \cos(\Delta\phi))} \tag{5.1}$$

where $\Delta\phi$ is the angle between the lepton and $E_{\mathrm{T}}^{\mathrm{miss}}$ in the transverse plane.

Visible mass refers to the invariant mass of the visible decay products in the detector. For example $m_{\mathrm{vis}}^{\mathrm{lep\,had}}$ refers to the visible mass of the lepton-τ_{had} system, and $m_{\mathrm{vis}}^{\mathrm{jets}}$ denotes the invariant mass of all jets in the event.

Missing Mass Calculator The visible mass of the tau decay products is not expected to match the Higgs boson mass, because the neutrinos involved in the decay carry away momentum. The missing transverse momentum of all neutrinos in the event is quantified as the missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$. The *missing mass calculator* [26] performs a likelihood scan that takes the momentum and angular information of the lepton, the tau and the missing transverse energy into account. This method yields the most probable mass m_{MMC} of the di-tau system, *i.e.*, the mass of the Higgs boson system.

Total transverse momentum $p_{\mathrm{T}}^{\mathrm{total}}$ is the magnitude of the vectorial sum of the transverse momenta of all visible decay products of the taus, the two leading jets, and $E_{\mathrm{T}}^{\mathrm{miss}}$, *i.e.*,

$$p_{\mathrm{T}}^{\mathrm{total}} = \left|\boldsymbol{p}_{\mathrm{T}}^{\mathrm{lep}} + \boldsymbol{p}_{\mathrm{T}}^{\mathrm{had}} + \boldsymbol{p}_{\mathrm{T}}^{j_0} + \boldsymbol{p}_{\mathrm{T}}^{j_1} + \boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}\right|. \tag{5.2}$$

Hadron Collider Moments refers to of eight variables, which capture the event topology [27]. The hadron collider moments are based on the Fox-Wolfram moments. The hadron collider moments h_i with $i = 1, \ldots 8$ are invariant under rotation around the beam and boosts parallel to the beam.

Transverse momentum of the Higgs boson p_{T}^{H} is defined as the magnitude of the vector sum of the transverse momentum of all visible decay products of the taus and $E_{\mathrm{T}}^{\mathrm{miss}}$

$$p_{\mathrm{T}}^{H} = \left|\boldsymbol{p}_{\mathrm{T}}^{\mathrm{lep}} + \boldsymbol{p}_{\mathrm{T}}^{\mathrm{had}} + \boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}\right|. \tag{5.3}$$

Scalar sum of transverse momentum $\sum p_{\mathrm{T}}$ is the scalar sum of the transverse momenta of the visible decay products of the tau and all jets. This variable gives information about the overall activity of the event and whether the Higgs boson system is boosted.

$E_{\mathrm{T}}^{\mathrm{miss}}\phi$ centrality measures the centrality of the $E_{\mathrm{T}}^{\mathrm{miss}}$ vector with respect to the position of the vector of the lepton and the tau in the transverse plane [15]. It is defined as

$$E_{\mathrm{T}}^{\mathrm{miss}}\phi\,\mathrm{centrality} = \frac{r+s}{\sqrt{r^2+s^2}}, \tag{5.4}$$

where

$$r = \frac{\sin(\phi^{\mathrm{miss}} - \phi^{\mathrm{had}})}{\sin(\phi^{\mathrm{lep}} - \phi^{\mathrm{had}})} \quad \text{and} \quad s = \frac{\sin(\phi^{\mathrm{lep}} - \phi^{\mathrm{miss}})}{\sin(\phi^{\mathrm{lep}} - \phi^{\mathrm{had}})}. \tag{5.5}$$

The centrality takes values between $\pm\sqrt{2}$. If $E_{\mathrm{T}}^{\mathrm{miss}}$ is perfectly central between the lepton and the tau, the centrality takes the value $\sqrt{2}$.

Lepton η centrality measures the centrality in pseudorapidity η of the lepton with respect to the leading j_0 and sub-leading jet j_1 [15]. It is defined as

$$\ell\,\eta\,\mathrm{centrality} = \exp\left(-\frac{4}{(\eta_{j_0}-\eta_{j_1})^2}\left(\eta_\ell - \frac{\eta_{j_0}+\eta_{j_1}}{2}\right)^2\right) \tag{5.6}$$

and takes the value 1 if the lepton is perfectly central between the two jets.

Collinear approximation is an alternative approach to reconstruct the mass of the Higgs boson system. It is, however, inferior to the missing mass calculation in terms of the resolution of the reconstructed mass [26]. The collinear approximation is that the momenta of the neutrino(s) and the visible decay products of the tau point to the same direction. By considering the missing transverse momentum $E_{\mathrm{T}}^{\mathrm{miss}}$, one is able estimate the momentum of the neutrinos. The variables x_0^{collin} and x_1^{collin} quantify the fraction of momentum that is carried by the visible decay products of the tau compared to the total momentum of the tau (visible products and invisible products).

5.3 Background and Signal Model

The background and signal model is based on Monte Carlo simulation with the exception of the fake background estimation which includes data-driven methods. The analysis uses Monte Carlo from the ATLAS mc15 production. Table 5.2 lists the samples used in this analysis.

To validate the background model, several control regions (CR) with negligible signal contribution are defined to check the agreement between the background estimation and data. The control regions are defined in Sec. 5.7.

Additionally a private Monte Carlo production for the signal in the multivariate analysis is used. Details about this private production are given in Sec. 6.4.

5.3.1 $Z \to \tau\tau$ Background Estimation

The background from $Z \to \tau\tau$ events represents one of the most important backgrounds in the $H \to \tau\tau$ analysis. In contrast to the procedure in Run 1, where

Table 5.2: Summary of the Monte Carlo samples, their generators, and the corresponding cross sections and branching ratios. The leptons in the diboson decays include leptons from all generations, *i.e.*, $\ell = e, \mu, \tau$.

Process	Generator	$\sigma \cdot \mathrm{BR}\,/\,\mathrm{pb}$
ggF $H \to \tau_{\mathrm{lep}}\tau_{\mathrm{had}}$	Powheg + Pythia8	1.262
VBF $H \to \tau_{\mathrm{lep}}\tau_{\mathrm{had}}$	Powheg + Pythia8	0.1079
$Z \to ee$ ($m_{ee} > 40\,\mathrm{GeV}$)	Madgraph + Pythia8	2111
$Z \to \mu\mu$ ($m_{\mu\mu} > 40\,\mathrm{GeV}$)	Madgraph + Pythia8	2104
$Z \to \tau\tau$ ($m_{\tau\tau} > 40\,\mathrm{GeV}$)	Madgraph + Pythia8	2097
$Z \to ee$ ($10\,\mathrm{GeV} < m_{ee} < 40\,\mathrm{GeV}$)	Madgraph + Pythia8	3399
$Z \to \mu\mu$ ($10\,\mathrm{GeV} < m_{\mu\mu} < 40\,\mathrm{GeV}$)	Madgraph + Pythia8	3253
$Z \to \tau\tau$ ($10\,\mathrm{GeV} < m_{\tau\tau} < 40\,\mathrm{GeV}$)	Madgraph + Pythia8	3074
VBF $Z \to ee$	Sherpa 2.1	2.545
VBF $Z \to \mu\mu$	Sherpa 2.1	2.538
VBF $Z \to \tau\tau$	Sherpa 2.1	2.541
$W \to e\nu$	Madgraph + Pythia8	20099
$W \to \mu\nu$	Madgraph + Pythia8	20094
$W \to \tau\nu$	Madgraph + Pythia8	20081
$t\bar{t}$	Powheg + Pythia6	832.9
Wt	Powheg + Pythia6	71.67
Single top (t-channel)	Powheg + Pythia6	70.43
Single top (s-channel)	Powheg + Pythia6	3.350
$VV \to \ell\ell\ell\ell$	Sherpa 2.1	11.65
$VV \to \ell\ell\ell\nu$	Sherpa 2.1	11.88
$VV \to \ell\ell\nu\nu$	Sherpa 2.1	12.75
$WW \to \ell\nu qq$	Sherpa 2.1	45.31
$WZ \to \ell\nu qq$	Sherpa 2.1	10.47
$WZ \to qq\ell\ell$	Sherpa 2.1	3.117
$WZ \to qq\nu\nu$	Sherpa 2.1	6.166
$ZZ \to qq\ell\ell$	Sherpa 2.1	2.146
$ZZ \to qq\nu\nu$	Sherpa 2.1	4.224

embedding was used to model $Z \to \tau\tau$ events based on $Z \to \mu\mu$ events from data, this analysis relies on Monte Carlo generators.

The $Z \to \tau\tau$ background has the same detector signature as the signal process. The difference, however, is the invariant mass of the di-tau system. The background from $Z \to \tau\tau$ events is expected to peak at around $m_Z = 91\,\mathrm{GeV}$, whereas the Higgs boson signal is expected to peak at $m_H = 125\,\mathrm{GeV}$.

5.3.2 $Z \to \ell\ell$ Background Estimation

The background originating from $Z \to \ell\ell$ decays (where $\ell = e, \mu$) is taken from Monte Carlo simulation. The dominant contribution from this background process enters the analysis, if one lepton from $Z \to \ell\ell$ is identified as the visible leptonic decay product of a tau and another object is misidentified as a hadronic tau. In case of $Z \to ee$, one electron can be misidentified as a τ_{had}.

5.3.3 Top Background Estimation

The top background has only a minor contribution in this analysis. Only approximately 3% of the events in the final signal regions are due to processes involving top quarks. The background model for these processes is taken directly from the Monte Carlo simulation.

This process affects the analysis, since top quarks decay via a W boson, similarly to the τ decay. The main contribution comes from $t\bar{t} \to bW^+\bar{b}W^-$ processes, where one W boson decays to leptons ($e\nu_e$ or $\mu\nu_\mu$) and the other to quarks. The hadronically decaying W mimics a τ_{had}. A minor contribution comes from Wt events, due to its low cross section, see Tab. 5.2. These background processes can be rejected by vetoing b-tagged events. Since the real W from the top decay manifests itself in large transverse mass, the rejection of this background can be enhanced by requiring $m_{\mathrm{T}} < 70\,\mathrm{GeV}$.

5.3.4 VV Background Estimation

The diboson background processes have only a tiny contribution of about 1% in the signal regions of this analysis. The background model for this background process is taken directly from the Monte Carlo simulation. There is no specific control region for this background, since this process has only a marginal contribution.

5.3.5 Data-Driven Fake Background Estimation

The background events where a jet has been misidentified as a τ_{had} are not taken from Monte Carlo because the Monte Carlo simulation is not reliable in the context of τ_{had} misidentification. The data-driven fake factor method is used to model events where a jet is misidentified as a τ_{had}, commonly referred to as a jet faking a τ, or simply fake taus. The development, implementation and validation for the fake factor method does not represent my own work and is not part of this thesis. However, this method is presented here because it is a vital part of the analysis.

To exclusively use the fake factor method for the fake background, events from the other background MC samples are removed, if the MC truth information indicates that the reconstructed τ_{had} was not a true τ, but rather a QCD jet. There are other possible sources of misidentification, for example, an electron can fake a τ_{had}. These cases have only a minor contribution in the signal regions and are not modeled with the fake factor method. Their background estimation is taken directly from Monte Carlo. Events with a fake τ_{had} come mainly from QCD, Z+jets, W+jets and $t\bar{t}$ processes.

For the fake factor method, taus failing the identification criteria are retained in anti-τ versions of the categories and regions of the analysis. For this, only the leading τ candidate is considered. All other cuts are left in place for the anti-τ regions. The fake rates are different for jets initiated by a gluon or by a quark. To keep the quark-gluon-ratio close to the ratio in the signal regions, a cut-off on the jet BDT score of 0.35 has been introduced for the events with a tau failing the identification.

The fake background is modeled by events from data where the tau fails the identification criteria. For this the data events from the anti-τ regions are added as fake background to the signal regions weighted by the fake factor f. To remove the contribution where the tau is not faked by a jet (denoted by $j \not\to \tau$), events from MC in the anti-τ regions, where the tau is not faked by a jet, are also added to the signal regions weighted by the negative fake factor $-f$. The yield $N_{\mathrm{fake}}^{\mathrm{SR}}$ of the fake background in the signal regions (SR) can be summarized by

$$N_{\mathrm{fake}}^{\mathrm{SR}} = (N_{\mathrm{data}}^{\text{anti-}\tau,\mathrm{SR}} - N_{\mathrm{MC},j\not\to\tau}^{\text{anti-}\tau,\mathrm{SR}}) \cdot f. \tag{5.7}$$

The fake factor f itself is calculated separately for each category and binned in the transverse momentum of the tau candidate $p_{\mathrm{T}}^{\mathrm{had}}$ and its number of tracks $n_{\mathrm{tracks}}^{\mathrm{had}}$, *i.e.*, $f = f(p_{\mathrm{T}}^{\mathrm{had}}, n_{\mathrm{tracks}}^{\mathrm{had}})$. The fake factor is calculated from a sum of process-specific fake factors f_i weighted by the expected relative yield contribution R_i in the anti-τ regions:

$$f = R_W f_W + R_Z f_Z + R_{\mathrm{Top}} f_{\mathrm{Top}} + R_{\mathrm{QCD}} f_{\mathrm{QCD}} \tag{5.8}$$

The relative yields R_i for physics process i are derived in the anti-τ signal region considering only events where a jet fakes a tau. Specifically, R_i is the ratio of the Monte Carlo yield of process i over the total yield. The total yield is estimated by the data yield in the anti-τ region reduced by the Monte Carlo yield in the anti-τ region where the tau is not faked by jet. Formulaically this reads

$$R_i = \frac{N_{i,MC,j\to\tau}^{\text{anti-}\tau,SR}}{N_{\mathrm{data}}^{\text{anti-}\tau,\mathrm{SR}} - N_{\mathrm{MC},j\not\to\tau}^{\text{anti-}\tau,\mathrm{SR}}}. \tag{5.9}$$

This procedure is used for the physics processes $i = W, Z, \mathrm{Top}$. The contribution from QCD processes is difficult to infer from MC, therefore the condition $\sum_i R_i = 1$ can be used to determine R_{QCD}. The relative fake yield of QCD processes is defined as

$$R_{\mathrm{QCD}} = 1 - \sum_{i \neq \mathrm{QCD}} R_i. \tag{5.10}$$

The individual fake factors f_i for physics process i are calculated in the individual control regions CR_i as the yield ratio of data events that fail the tau identification over events that pass the identification. The control regions of this analysis are defined in Sec. 5.7. The contribution from other processes (denoted by "not i") in the control regions are estimated by Monte Carlo and subtracted from the data yields. Similarly the contribution from events where the tau is not faked by a jet is also estimated by Monte Carlo and subtracted from the data yield. The fake factor calculation then reads

$$f_i = \frac{N^{\mathrm{CR}_i}_{\mathrm{data}} - N^{\mathrm{CR}_i}_{\mathrm{MC,not}\,i} - N^{\mathrm{CR}_i}_{i,\mathrm{MC},j\not\to\tau}}{N^{\text{anti-}\tau,\mathrm{CR}_i}_{\mathrm{data}} - N^{\text{anti-}\tau,\mathrm{CR}_i}_{\mathrm{MC,not}\,i} - N^{\text{anti-}\tau,\mathrm{CR}_i}_{i,\mathrm{MC},j\not\to\tau}}. \tag{5.11}$$

Various closure tests and validation studies have been carried out to confirm this method. The description of these studies is beyond the scope of this thesis.

5.3.6 Signal Model

The main signal model is built from inclusive vector boson fusion and gluon fusion Monte Carlo samples. Other signal processes have been investigated, but have been discarded due to their small contribution. For dedicated multivariate analysis studies, filtered ggF $H \to \tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ Monte Carlo samples have been produced, which is discussed in detail in Sec. 6.4. A Higgs boson mass of $m_H = 125\,\mathrm{GeV}$ is assumed throughout.

5.4 Preselection

The preselection defined in the following is shared among the two analysis variants. The goal of the preselection is to reject events from various background processes without exploiting the event topology specific to the Higgs boson production mechanism.

The following conditions are imposed on all events.

1. There must be exactly one lepton (electron or muon) in the event, which

   (a) satisfies the gradient isolation criteria,

   (b) passes the medium identification working point and

   (c) exceeds the offline $p_{\mathrm{T}}^{\mathrm{lep}}$ threshold of 25 GeV for electrons and 21 GeV (25.2 GeV) for muons in 2015 (2016).

   The lepton stems from the leptonically decaying tau. The isolation criteria suppresses background from QCD processes. The requirement of exactly one lepton rejects $Z \to \ell\ell$ and diboson background events.

2. There must be at least one τ_{had} in the event. The following conditions are only imposed on the τ_{had} candidate with the highest transverse momentum. Other τ_{had} in the event are ignored. The τ_{had} must

   (a) be of medium quality with an absolute charge of 1,

   (b) be in the pseudorapidity range $|\eta^{\mathrm{had}}| < 2.4$ (excluding the region with support infrastructure between 1.37 and 1.52) and

(c) exceed the offline threshold of $p_{\mathrm{T}}^{\mathrm{had}} > 20\,\mathrm{GeV}$.

3. The hadronic and leptonic tau candidates are required to have opposite charge, which suppresses events where either object is faked and/or originates from a different source.

4. A b-veto criteria is imposed. This means the event is rejected if there is at least one b-tagged jet with $p_{\mathrm{T}}^{j} > 20\,\mathrm{GeV}$ and $|\eta^{j}| < 2.5$. The flavor tagging uses the mv2c10 algorithm with an efficiency working point of 85%. The b-veto efficiently rejects top background processes.

5. The transverse mass m_{T} must satisfy $m_{\mathrm{T}} < 70\,\mathrm{GeV}$. This requirement suppresses backgrounds from processes with a real W boson decay.

Figure 5.1 shows selected distributions of important kinematic variables after applying the preselection requirements. The error bands correspond to the combination of the statistical and systematic uncertainty of the background model. Systematic uncertainties are discussed in Sec. 7.1. The distributions show mostly good agreement between the background expectation and the measured data. The distributions for p_{T}^{H}, $p_{\mathrm{T}}^{j_0}$, $E_{\mathrm{T}}^{\mathrm{miss}}$ show a slope in the ratio plot. This gives rise to further studies, to investigate the source of this problem, which are, however, beyond the scope of this thesis.

5.5 Cut-Based Signal Regions

The two dominant Higgs boson production mechanisms for this analysis are vector boson fusion and gluon fusion. All events passing the preselection stage are split into two categories: VBF and Boosted. The category definitions exploit different event topologies to separate the two production mechanisms. Events from vector boson fusion usually feature at least two jets with large η separation. Events from gluon fusion can be enhanced by requiring large transverse momenta of the reconstructed Higgs boson. This basic strategy is the same for the CBA and the MVA. However, the exact event category definitions differ slightly for the two analysis methods. For the cut-based analysis, the two categories are further split into sub-categories to maximize the sensitivity of the analysis. This section describes the signal regions for the CBA.

The VBF inclusive category for the CBA consists of all events passing the following requirements.

1. There must be at least two jets in the event.

2. The transverse momentum of the leading jet must satisfy $p_{\mathrm{T}}^{j_0} > 40\,\mathrm{GeV}$.

3. The transverse momentum of the sub-leading jet must satisfy $p_{\mathrm{T}}^{j_1} > 30\,\mathrm{GeV}$.

4. The two leading jets must be separated by $\Delta\eta^{jj} > 3.0$ and must be in opposite hemispheres of the detector, *i.e.*, $\eta^{j_0} \cdot \eta^{j_1} < 0$.

5. The invariant mass of the two leading jets m^{jj} must be greater than 300 GeV.

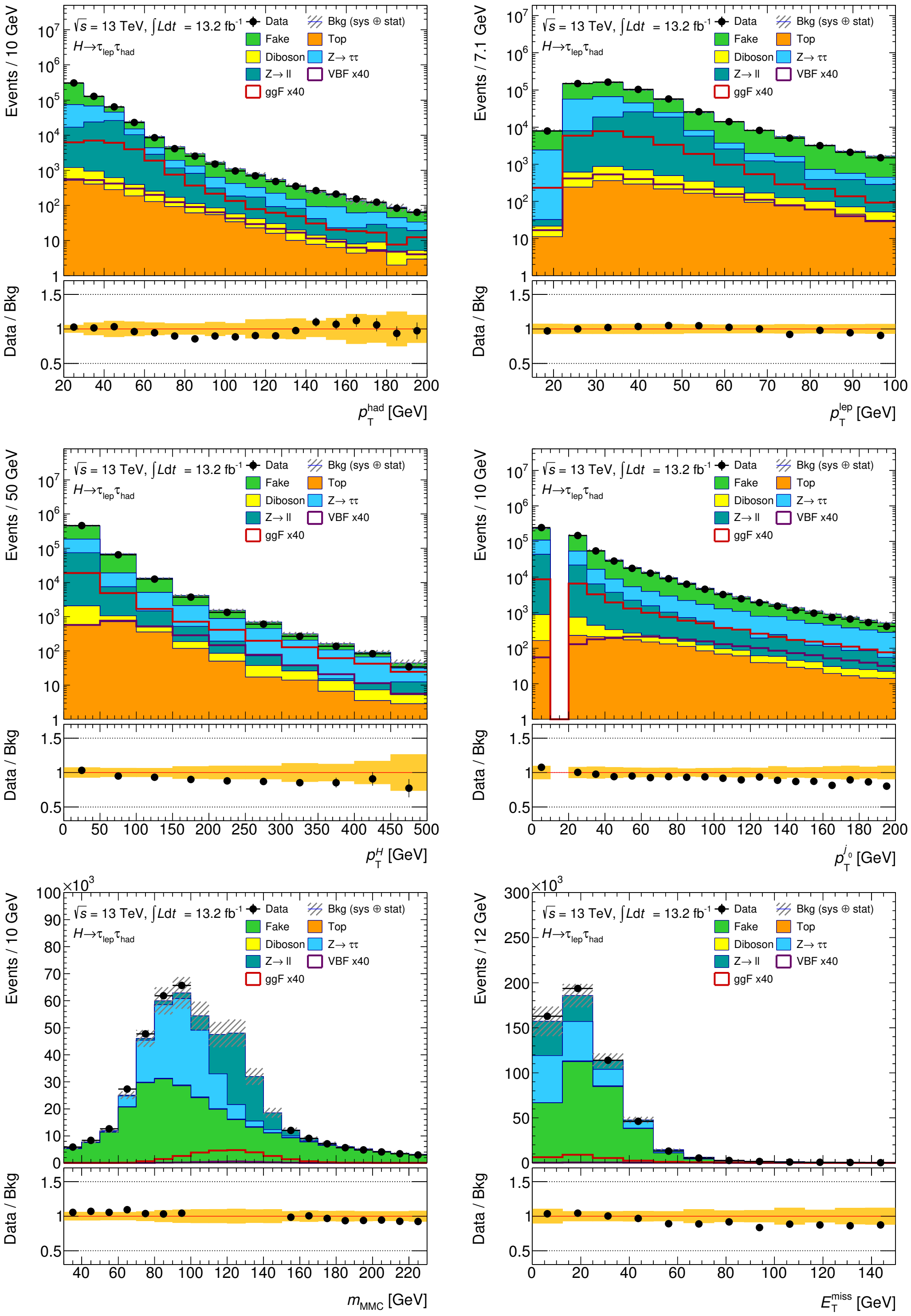

Figure 5.1: Selected distributions at preselection level. The error bands include statistical and systematic uncertainties. The top row shows the transverse momentum of the tau p_T^{had} (left) and the lepton p_T^{lep} (right). The middle row shows the transverse momentum of the Higgs boson p_T^H (left) and the leading jet $p_T^{j_0}$ (right), where entries with $p_T^{j_0} = 0$ indicate that no jet with $p_T^{j_0} \geq 20\,\mathrm{GeV}$ was present in the event. The bottom row shows the mass m_{MMC} (left) and the missing transverse energy E_T^{miss} (right). For the bottom left plot, no data entries are shown in the most signal-sensitive region $100\,\mathrm{GeV} < m_{\mathrm{MMC}} < 150\,\mathrm{GeV}$.

6. The pseudorapidity of the lepton and the hadronic tau must be between the pseudorapidity of the leading and the sub-leading jet. This requirement is referred to as η centrality.

7. The missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$ must be greater than 20 GeV.

8. The pseudorapidity difference $\Delta\eta^{\mathrm{lep\,had}}$ of the lepton and the hadronically decaying tau must be less than 1.5.

9. The angular difference in the η-ϕ plane $\Delta R^{\mathrm{lep\,had}}$ of the lepton and the hadronically decaying tau must not exceed 3.0.

All events that fulfill the VBF inclusive requirements are further split into two subcategories VBF tight and VBF loose. Events in the VBF tight category must satisfy the following requirements.

1. The invariant mass of the two leading jets m^{jj} must exceed 500 GeV.

2. The transverse momentum p_{T}^{H} of the reconstructed Higgs must be greater than 100 GeV.

3. The visible mass m_{vis} must be greater than 40 GeV.

4. The transverse momentum $p_{\mathrm{T}}^{\mathrm{had}}$ of the hadronic tau must be greater than 30 GeV.

The VBF loose region consists of all events passing the inclusive VBF criteria but failing the tight selection.

The Boosted inclusive category is defined by the following requirements.

1. The event must satisfy the preselection requirements, but fail the selection criteria of the inclusive VBF category.

2. The transverse momentum p_{T}^{H} of the reconstructed Higgs must be greater than 100 GeV.

3. The missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$ must be greater than 20 GeV.

4. The transverse momentum $p_{\mathrm{T}}^{\mathrm{had}}$ of the hadronic tau must be greater than 30 GeV.

5. The pseudorapidity difference $\Delta\eta^{\mathrm{lep\,had}}$ of the lepton and the hadronically decaying tau must be less than 1.5.

6. The angular difference $\Delta R^{\mathrm{lep\,had}}$ of the lepton and the hadronically decaying tau must not exceed 2.5.

The Boosted category is subsequently split into a Boosted high and Boosted low sub-category depending on the transverse momentum p_{T}^{H} of the Higgs. Boosted high is defined by the following requirements:

1. The transverse momentum p_{T}^{H} of the Higgs must be greater than 140 GeV.

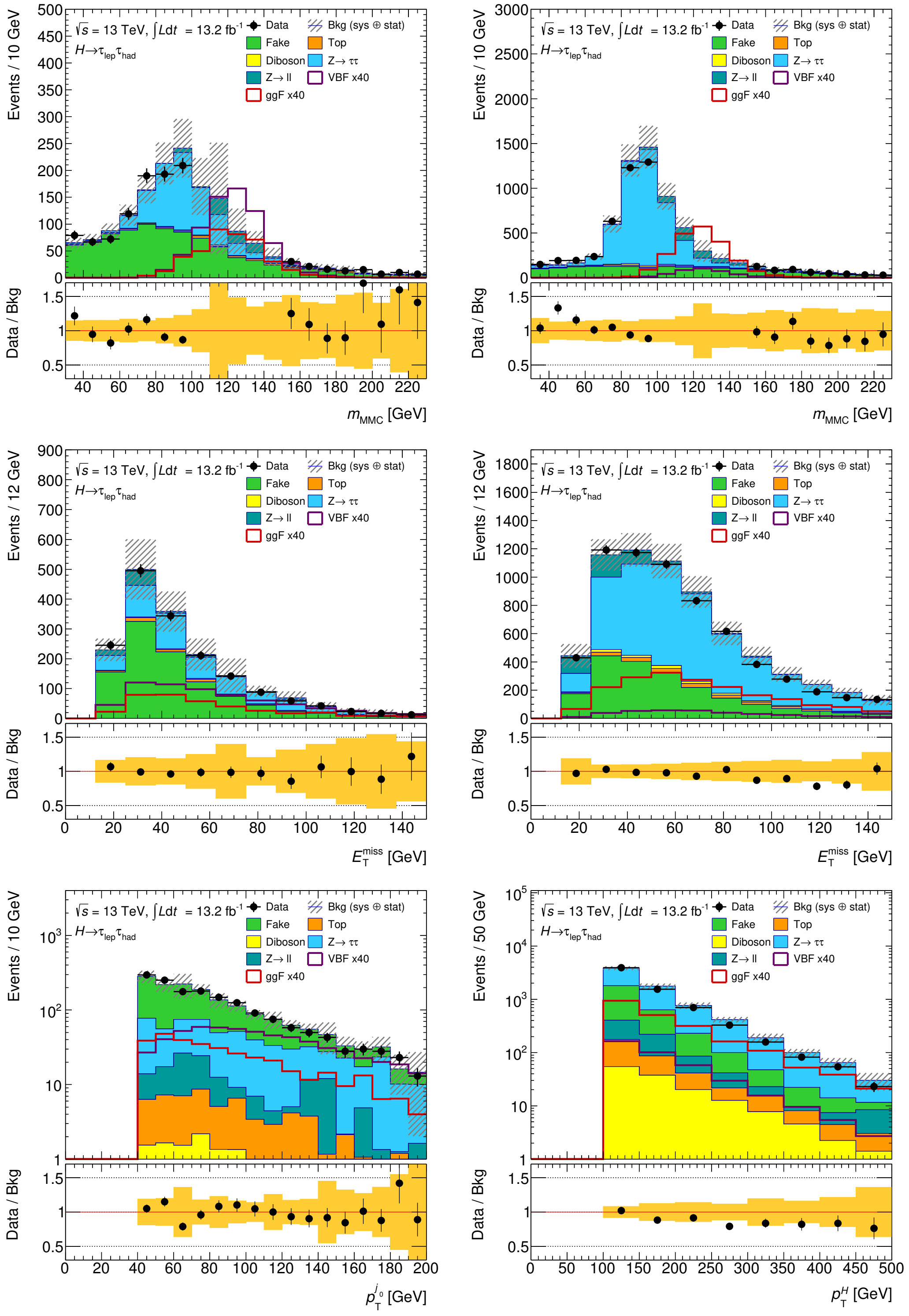

Figure 5.2: Selected distributions in the VBF inclusive (left) and Boosted inclusive categories (right) of the cut-based analysis. The error bands include statistical and systematic uncertainties. The top row shows the mass m_{MMC}, blinded in the signal-sensitive region. The middle row shows the missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$. The bottom row shows the transverse momentum of the leading jet $p_{\mathrm{T}}^{j_0}$ for the VBF category (left) and the transverse momentum of the Higgs boson p_{T}^{H} for the Boosted category (right).

2. The angle $\Delta R^{\mathrm{lep\,had}}$ between the lepton and the hadronically decaying tau must be less than 1.5.

All events that pass the Boosted selection criteria, but fail the Boosted high requirements, constitute the Boosted low category.

Figure 5.2 shows selected distributions of important kinematic variables in the VBF and Boosted inclusive regions of the cut-based analysis. The distributions show good agreement between the background expectation and the measured data. However, the distribution of p_{T}^{H} shows a slope in the ratio plots. This feature was already observed at the preselection level.

5.6 Multivariate Signal Regions

Following the idea of the CBA categorization, two similar categories for the multivariate analysis are defined. The categories are looser[1] than the CBA categories to retain Monte Carlo statistics, which is essential for MVA training. The analysis follows the definitions from Run 1 [15].

The VBF category for the MVA is defined by all events passing the preselection and the following requirements:

1. There must be at least two jets in the event.
2. The transverse momentum of the leading jet must satisfy $p_{\mathrm{T}}^{j_0} > 50\,\mathrm{GeV}$.
3. The transverse momentum of the sub-leading jet must satisfy $p_{\mathrm{T}}^{j_1} > 40\,\mathrm{GeV}$.
4. The two leading jets must be separated by $\Delta\eta^{jj} > 3.0$.
5. The visible mass m_{vis} must be greater than 40 GeV.

The Boosted category consists of all events passing the preselection and failing the VBF requirements that also satisfy that

1. the transverse momentum p_{T}^{H} of the Higgs boson must be greater than 100 GeV.

The two MVA categories are not split any further, in contrast to the sub-categorization for the CBA. Introducing sub-categories in the multivariate analysis would lead to regions with less statistics, which would have negative effects on the performance of a classifier. The idea behind the MVA is, that the classifier learns how to combine kinematic variables to enhance the separation between background and signal.

Figure 5.3 shows selected distributions of important kinematic variables in the VBF and Boosted regions of the multivariate analysis. The distributions show good agreement between the background expectation and the measured data. However, the distribution of p_{T}^{H} shows a slope in the ratio plots, which was already observed at preselection level and in the cut-based analysis.

The event selection for the cut-based analysis and multivariate analysis is summarized in Tab. 5.3.

[1] With the exception of the p_T^j of the leading jet and sub-leading jet. In the cut-based analysis the threshold is 40 GeV (30 GeV) whereas the threshold in the multivariate analysis is 50 GeV (40 GeV) for the leading (sub-leading) jet.

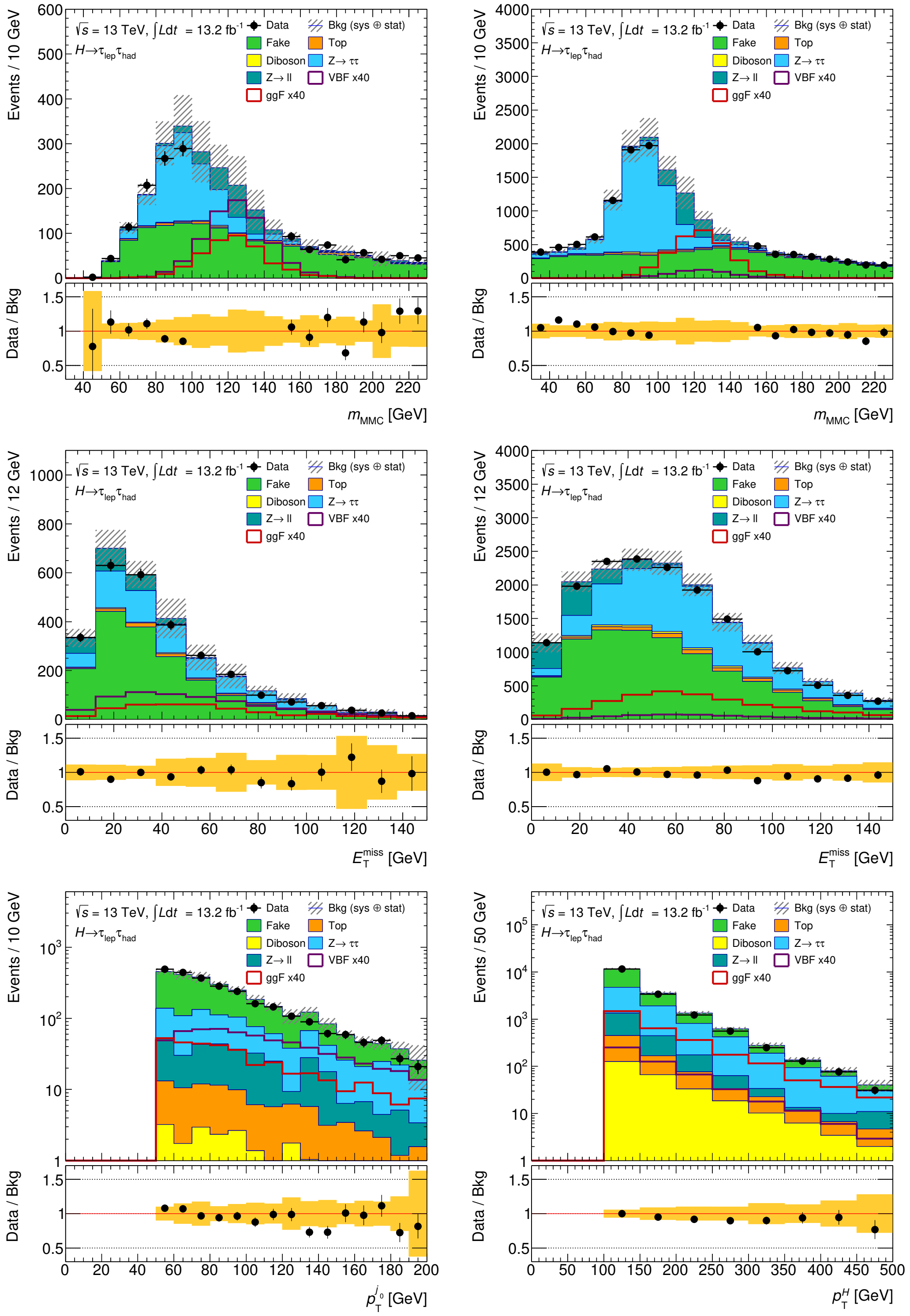

Figure 5.3: Selected distributions in the VBF (left) and Boosted categories (right) of the multivariate analysis. The error bands include statistical and systematic uncertainties. The top row shows the mass m_{MMC}, blinded in the signal-sensitive region. The middle row shows the missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$. The bottom row shows the transverse momentum of the leading jet $p_{\mathrm{T}}^{j_0}$ for the VBF category (left) and the transverse momentum of the Higgs boson p_{T}^{H} for the Boosted category (right).

Table 5.3: Summary of the event selection of the cut-based analysis and the multivariate analysis. The preselection criteria are shared between cut-based analysis and multivariate analysis.

Preselection			
• lepton: $n^{\mathrm{lep}} = 1$, gradient isolation, medium id. $p_{\mathrm{T}}^{e} > 25\,\mathrm{GeV}$ or $p_{\mathrm{T}}^{\mu} > 21\,\mathrm{GeV}$ $(25.2\,\mathrm{GeV})$ in 2015 (2016) • τ_{had}: $n^{\mathrm{had}} \geq 1$, medium id., $\lvert q \rvert = 1$, $\lvert \eta^{\mathrm{had}} \rvert < 2.4$, $p_{\mathrm{T}}^{\mathrm{had}} > 20\,\mathrm{GeV}$ • $q^{\mathrm{had}} \cdot q^{\mathrm{lep}} < 0$ • b-veto • $m_{\mathrm{T}} < 70\,\mathrm{GeV}$			
CBA			
VBF inclusive		Boosted (Bst.) inclusive	
• pass preselection • $p_{\mathrm{T}}^{j_0} > 40\,\mathrm{GeV}$ • $p_{\mathrm{T}}^{j_1} > 30\,\mathrm{GeV}$ • $\Delta\eta^{jj} > 3.0$ • $\eta^{j_0} \cdot \eta^{j_1} < 0$ • $m^{jj} > 300\,\mathrm{GeV}$ • η centrality • $E_{\mathrm{T}}^{\mathrm{miss}} > 20\,\mathrm{GeV}$ • $\Delta\eta^{\mathrm{lep\,had}} < 1.5$ • $\Delta R^{\mathrm{lep\,had}} < 3.0$		• pass preselection • fail VBF incl. • $p_{\mathrm{T}}^{H} > 100\,\mathrm{GeV}$ • $E_{\mathrm{T}}^{\mathrm{miss}} > 20\,\mathrm{GeV}$ • $p_{\mathrm{T}}^{\mathrm{had}} > 30\,\mathrm{GeV}$ • $\Delta\eta^{\mathrm{lep\,had}} < 1.5$ • $\Delta R^{\mathrm{lep\,had}} < 2.5$	
VBF tight	VBF loose	Boosted high	Boosted low
• pass VBF incl. • $m^{jj} > 500\,\mathrm{GeV}$ • $p_{\mathrm{T}}^{H} > 100\,\mathrm{GeV}$ • $m_{\mathrm{vis}} > 40\,\mathrm{GeV}$ • $p_{\mathrm{T}}^{\mathrm{had}} > 30\,\mathrm{GeV}$	• pass VBF incl. • fail VBF tight	• pass Bst. incl. • $p_{\mathrm{T}}^{H} > 140\,\mathrm{GeV}$ • $\Delta R^{\mathrm{lep\,had}} < 1.5$	• pass Bst. incl. • fail Bst. high
MVA			
VBF		Boosted	
• pass preselection • $p_{\mathrm{T}}^{j_0} > 50\,\mathrm{GeV}$ • $p_{\mathrm{T}}^{j_1} > 40\,\mathrm{GeV}$ • $\Delta\eta^{jj} > 3.0$ • $m_{\mathrm{vis}} > 40\,\mathrm{GeV}$		• pass preselection • fail VBF • $p_{\mathrm{T}}^{H} > 100\,\mathrm{GeV}$	

Table 5.4: Definition of the control regions. The control regions are enriched in the processes listed in the third column. The fourth column shows the approximate purity at the preselection level.

Region	Definition	Enriched in	Purity
Top CR	invert b-veto and invert m_{T} cut	top	59%
W CR	invert m_{T} cut	fake	91%
Z CR	require di-lepton system (see Sec. 5.7)	fake	96%
QCD CR	invert lepton isolation criteria	fake	95%

5.7 Control Regions

In order to validate the background model, various control regions with negligible signal contribution have been defined. The control regions are intended to be enriched with one background process in order to verify the modeling of this physics process. To prevent a bias in the control regions, the regions are chosen to be close to the signal region. The control regions are defined by inverting (in most cases) only one cut, which was introduced in the standard event selection to reject that type of background process. Table 5.4 lists the definitions of all the control regions used in the MVA and CBA analysis.

The control region for top background is defined by requiring a b-tagged jet and large transverse mass $m_{\mathrm{T}} > 70\,\mathrm{GeV}$. The top control region has a purity of roughly 59%. Studies have been performed to increase the purity by using different values for the m_{T} cut. The overall performance of the analysis, however, did not increase, so the purity optimization has not been used for this thesis in favor of a control region closer to the signal region.

Figure 5.4 shows distributions in the top control region. The top control region plays a crucial role in the analysis, since it is used in the likelihood fit to estimate the top normalization. The background model shows good agreement with data in the top control region.

The top region used to derive the fake factors is slightly different from the definition of the top control region above. In contrast to the definition in Tab. 5.4, the cut on the transverse mass has been changed to $m_{\mathrm{T}} > 40\,\mathrm{GeV}$.

The W control region is enriched in $W \to \ell\nu_\ell$ ($\ell = e, \mu, \tau$) decays. It is defined by inverting the m_{T} cut to $m_{\mathrm{T}} > 70\,\mathrm{GeV}$. The main contribution of W to the analysis comes from events where the W decays to light leptons (e, μ) and a jet fakes a τ_{had}. The content of the W control region is dominated by fakes, since events with a jet faking a τ_{had} is modeled with the fake factor method, see Sec. 5.3.5. Figure A.1 shows selected distributions in the W control region. The discrepancy between data and Monte Carlo in the W control region at preselection level, gives rise to further studies, which are beyond the scope of this thesis. This discrepancy, however, does not impact the analysis, since it seems to be not significant in the VBF and Boosted categories for the CBA and the MVA given the uncertainties of the background and the data.

The Z control region is enriched in $Z \to \ell\ell$ decays (where $\ell = e, \mu$). It is defined

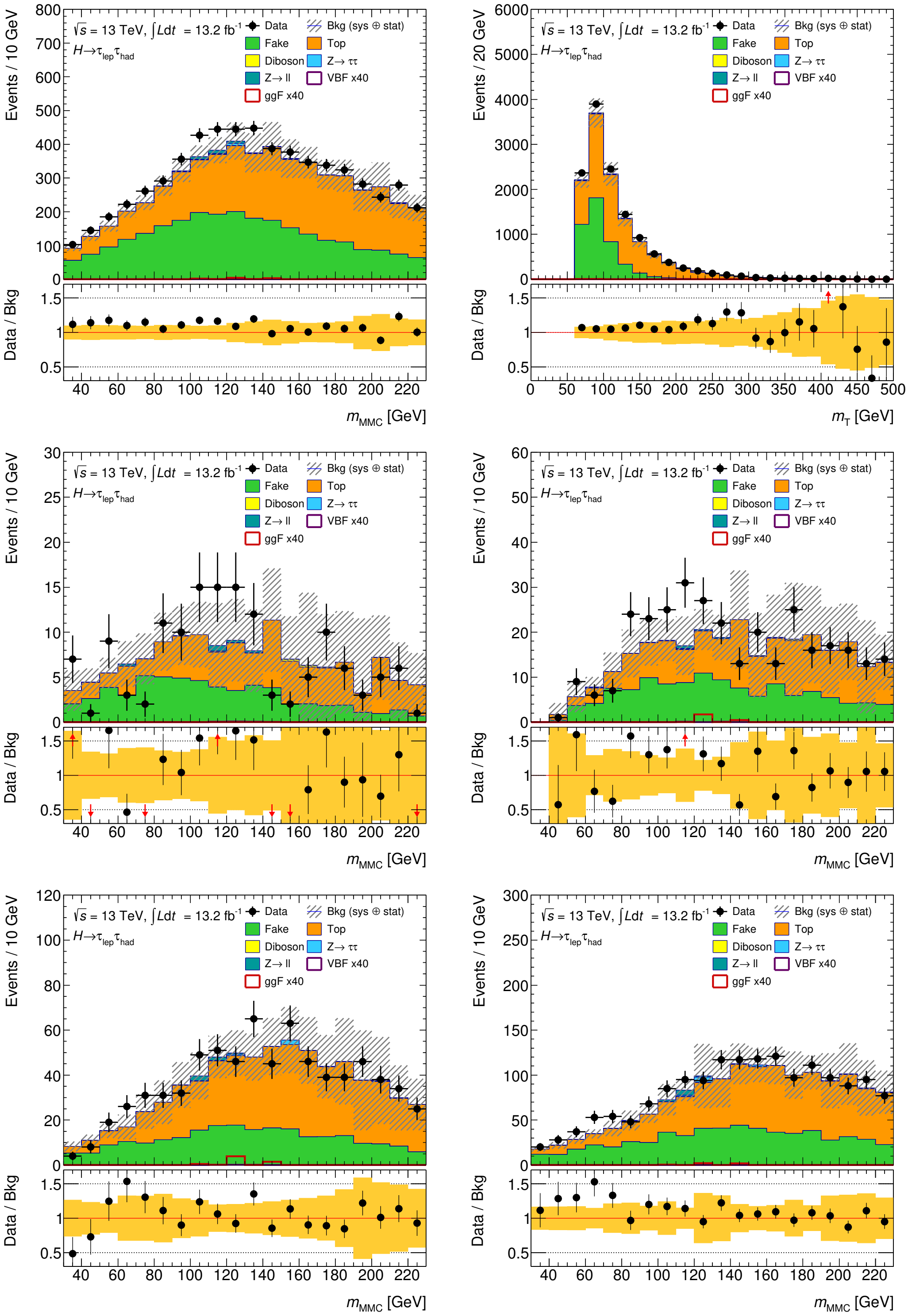

Figure 5.4: Selected distributions in the top control region. The error bands include statistical and systematic uncertainties. All plots show the mass m_{MMC}, except the top right plot which shows m_{T}. The top row shows the distributions after applying the preselection cut. The middle row shows the distributions for the VBF categories of the CBA (left) and MVA (right). The bottom row shows the distributions for the Boosted category of the CBA (left) and the MVA (right).

by requesting two same-flavor leptons of opposite charge. The invariant mass of the di-lepton system must be within the Z boson mass window between 61 GeV and 121 GeV. The conditions on the hadronic τ are not modified. The Z CR consists mostly of events modeled with the fake factor method. The reason for this is the requirement of two light leptons and a tau. The two light leptons typically come from Z decays, whereas the τ_{had} originates from a jet, which is misidentified as a τ_{had}. There are other scenarios, where an event can end up in this control region. It is, for example, possible to have a true τ_{had} and a jet faking an electron, but the above example is the most dominant contribution. Figure A.2 shows selected distributions in the Z control region. The background model shows good agreement with data in this control region.

The QCD control region is defined by inverting the lepton isolation criteria. This control region consists mostly of events where a QCD jet fakes a tau. Figure A.3 shows selected distributions in the QCD control region. The background model shows good agreement with data in this control region.

CHAPTER 6

Optimization of the Multivariate Analysis

This chapter comprises the main studies that have been performed during the course of this thesis. All the studies shown here are carried out separately for the VBF and Boosted categories unless otherwise noted. The first section introduces the main techniques and concepts of machine learning, which are used to optimize the parameters of boosted decision trees. Section 6.2 shows how the performance of a boosted decision tree is assessed in this thesis. The input variables of the boosted decision tree are chosen in Sec. 6.3. Section 6.4 discusses the issue of a lack of training statistics, which is addressed by a special filtered Monte Carlo production. Section 6.5 revisits the concept of k-fold cross validation introduced in Sec. 6.1 and studies its role in the analysis. The boosted decision tree training parameters are optimized in Sec. 6.6. The chapter closes with the validation of the selected BDTs in Sec. 6.7.

6.1 Introduction to Machine Learning

The multivariate analysis uses geometrically complex structures in the multi-dimensional phase space to separate signal and background events. Since it is difficult to observe or define these structures by hand, one usually employs machine learning techniques. Problems in machine learning or pattern recognition can be grouped into three classes: reenforcement learning, unsupervised learning, and supervised learning [28]. Reenforcement learning deals with problems where an algorithm has to find appropriate actions in a virtual environment to achieve a certain goal. In unsupervised, learning an algorithm is presented with a collection of observations and the algorithm has to group the observations based on their features. The third method, supervised learning, is best suited for the multivariate analysis at hand. In supervised learning in the context of particle physics, the algorithm is presented with Monte Carlo events of background and signal [19]. The algorithm learns what signal and background events *look* like. Since the algorithms in this context try to assign the class labels signal s and background b to unknown events, these types of algorithms are called *classifiers*. A classifier can be used on measured data, for which

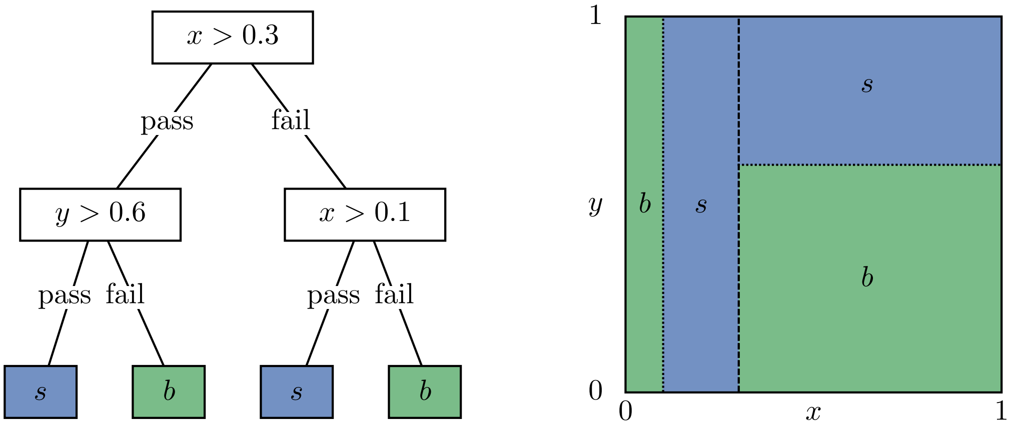

Figure 6.1: Structure of a decision tree (left) and the assignment of this decision tree in phase space (right). The nodes in the structure are represented by boxes. The terminal nodes are represented by colored boxes, the color depends on the class association. For simplicity the phase space is spanned by the two variables $x, y \in [0, 1]$. The split in phase space introduced by the first node is indicated by a dashed line, the splits introduced by the second layer of nodes is indicated by dotted lines.

the class association is unknown, to determine whether the event is more signal-like or background-like.

The field of machine learning hosts many different methods and algorithms to classify objects [28, 29]. Boosted decision trees have become popular especially in particle physics. This thesis will focus on the usage of a boosted decision tree. Its concept is described in Sec. 6.1.2.

6.1.1 Decision Trees

A decision tree is a classifier that splits the feature space, or the phase space in high energy physics, into rectangular regions[1]. The tree consists of nodes. Each node is annotated by a condition on a kinematic variable, similar to a cut in the cut-based analysis. Each internal node of the tree has two daughter nodes. The events that satisfy the node's condition are forwarded to the "pass" daughter node. The events that do not satisfy the condition are forwarded to the "fail" daughter node. The procedure is repeated until the event reaches a terminal node called *leaf node*. The structure of a decision tree is illustrated in Fig. 6.1. Each node splits the phase space in rectangular regions. The two sub-trees mounted to a given node act on two disjunct regions in phase space. The leaf nodes are annotated by the class labels signal (s) or background (b). If an event is passed along the tree and reaches a signal leaf node, the classifier considers the event as signal-like.

The operation of a decision tree can be illustrated by representing the phase space as a two dimensional surface. Then the whole area of the phase space would be covered with rectangles or unions of rectangles. Each point in phase space is labeled as either signal (s) or background (b). The phase space of a simple decision tree is sketched in Fig. 6.1.

[1] The consideration here is limited to binary decision trees.

The advantage of a decision tree is the simplicity of this classifier. The interpretation is directly apparent from the structure of the tree. A single path from the root node of the tree to a signal leaf node corresponds to a cut based analysis. Since a decision tree has many branches, it is more powerful than a cut-based analysis. The disadvantages come with the generation process of a decision tree. It is computationally difficult to optimize all cuts in the tree at the same time to minimize the misclassification error. Usually a simpler, greedy algorithm is used to construct a decision tree. The tree is built top down, *i.e.*, starting from the root node. At each node all input variables and all cut positions are scanned to see which cut maximizes the separation between signal and background in the daughter nodes. There are several methods to quantify the separation gain achieved by a single cut [29]. Throughout this thesis the decision trees are built by minimizing the gini index [30], which is defined as $p \cdot (1 - p)$, where p is the signal purity of the node. The background purity is then given by $(1 - p)$. A node which enhances background in one of its sub-nodes is as valuable as one which enhances signal. The gini index therefore has the advantage, that it is symmetric and does not prefer enhancing signal over enhancing background.

There are different methods to limit the size of the trees. One method is to impose a *maximal depth* of the trees, which means that all nodes at the specified depth are considered terminal nodes and do not feature any daughter nodes. Another method is to stop building a branch in the tree if only a fraction of the initial events is affected by the branch. This criteria is referred to as *minimal node size.*

Trees built with the algorithm outlined above are in general not optimal. The major disadvantage of decision trees using such a greedy algorithm is that the trees are unstable. Slight changes in the input dataset lead to completely different tree structures, since the differences propagate through the tree [29].

6.1.2 Boosting and Boosted Decision Trees

Boosting is one of the most powerful learning algorithms. The concept of Boosting is a procedure to combine other classifiers into a more powerful *committee* [29]. Boosting is neither limited to decision trees nor to classification problems, the discussion in this section, however, is limited to decision tree based classifiers. The idea of Boosting is to iteratively repeat the training of a single decision tree and increase the event weights for events which have been misclassified in the previous round. This way the construction of a single decision tree is forced to *focus* on events that are difficult to classify. The final decision of the *boosted decision tree* (BDT) is taken from a weighted average of the individual decision trees. The collection of decision trees is referred to as a committee.

Some notation has to be introduced to study the procedure of boosting in more detail. The set used to train the boosted decision tree consists of N events with event weights w_n. The n-th event is characterized by a vector $\boldsymbol{x}_n$ comprised of all kinematic variables of the event. The class labels for the training events are $t_n \in \{-1, +1\}$, where $t_n = -1$ $(+1)$ means that the n-th event is a background (signal) event. The committee consists of M decision trees. The prediction of the m-th decision tree in the committee when used on event n is $d_m(\boldsymbol{x}_n) \in \{-1, +1\}$.

The weighted average of the committee when used on event n is defined as

$$f(\boldsymbol{x}_n) = \sum_{m=1}^{M} \alpha_m d_m(\boldsymbol{x}_n). \tag{6.1}$$

The value of the function $f(\boldsymbol{x}_n)$ is the prediction of the boosted decision tree. The prediction of the BDT is called *output score* and due to the weighted sum, can take any values in the interval $[-1, 1]$. One approach [31] to understand boosting is to start with a loss function, for example the exponential loss function

$$L(t_n, f(\boldsymbol{x}_n)) = e^{-t_n f(\boldsymbol{x}_n)}, \tag{6.2}$$

which quantifies the penalty of a false prediction $f(\boldsymbol{x}_n) \neq t_n$ or the reward (small values of L) in case of correct prediction $f(\boldsymbol{x}_n) = t_n$. It is useful to define the classification error ϵ_m of the decision tree d_m as

$$\epsilon_m = \frac{\sum_{n=1}^{N} w_n \, I(t_n \neq d_m(\boldsymbol{x}_n))}{\sum_{n=1}^{N} w_n} \tag{6.3}$$

where $I(\text{cond.}) \in \{0, 1\}$ is an indicator function with $I = 1$ if and only if the condition is true. The committee is trained iteratively while minimizing the sum of losses

$$E = \sum_{n=1}^{N} L(t_n, f(\boldsymbol{x}_n)). \tag{6.4}$$

Since the weights of the events are updated iteratively, the weight of event n, when it is used with decision tree d_m, is $w_n^{(m)}$. Assume the committee already consists of $m-1$ fixed classifiers $d_1, \dots d_{m-1}$ with weights $\alpha_1, \dots, \alpha_{m-1}$. Adding a new decision tree d_m to the committee requires a new minimization of the sum of losses E. Since all preexisting classifiers are fixed, d_m and α_m are the only variable terms of the loss function. It can be shown [28] that this minimization of E with L as defined in Eq. (6.2) translates to the following three statements. Firstly, the minimization of E implies, that the weights of events, which were falsely classified by the previous tree d_{m-1} (*i.e.* $t_n \neq d_{m-1}(\boldsymbol{x}_n)$), should be updated according to

$$w_n^{(m-1)} \to w_n^{(m)} = w_n^{(m-1)} \cdot e^{\alpha_{m-1}}. \tag{6.5}$$

Secondly, it implies that the classification error ϵ_m should be minimized. This is the usual condition for the training of a classifier. Lastly, the minimization of E implies that the weight α_m of d_m in the committee is determined by

$$\alpha_m = \log\left(\frac{1 - \epsilon_m}{\epsilon_m}\right). \tag{6.6}$$

The usage of the exponential loss function in Eq. (6.2) leads to the simple Equations (6.5, 6.6) which are commonly known as *AdaBoosting.* Using different loss functions leads to different boosting properties. Depending on the loss function, the minimization of E does not necessarily lead to simple analytic functions for the adjustment of the event and decision tree weights α_m and $w_n^{(m)}$. In some cases numerical methods need to be applied [29]. An alternative boosting method is *gradient boosting.* The loss function of gradient boosting [30] is

$$L(t_n, f(\boldsymbol{x}_n)) = \log\left(1 + e^{-2t_n f(\boldsymbol{x}_n)}\right). \tag{6.7}$$

When the loss function of AdaBoosting is compared to the loss function of gradient boosting, it becomes apparent, that AdaBoosting has a larger penalty term for outliers compared to gradient boosting. AdaBoosting is therefore less robust to outliers [28]. All boosted decision trees in this thesis employ gradient boosting. The software package TMVA [30] is used throughout this thesis to train the boosted decision trees.

The classification power of the committee is increased by each decision tree that is added, because each new tree tries to correct the mistakes of the committee. The stability and robustness of the boosted decision tree can be increased by artificially slowing down the learning process of the committee. Let the committee consisting of m decision trees be denoted by f_m. A new parameter, the *shrinkage* $\nu \in (0, 1]$, is introduced to slow down the learning process. When decision tree d_m is added to the committee f_{m-1}, the new committee is formed according to

$$f_m(\boldsymbol{x}_n) = f_{m-1}(\boldsymbol{x}_n) + \nu\alpha_m d_m(\boldsymbol{x}_n). \tag{6.8}$$

6.1.3 Overtraining

When a boosted decision tree is trained, it tries to fit its prediction to the class labels in the training set. The boosted decision tree adapts to smaller and smaller features in the training observations with increasing complexity of the boosted decision tree, for example with increasing number of trees, increasing shrinkage values, or larger individual decision trees. At some point the features that the BDT is picking up are merely random fluctuations due to the limited amount of the training observations. The boosted decision tree will become perfectly adapted to the training observations with increasing BDT complexity.

If this trend extends to cases where the BDT is not able to correctly classify new observations, which are not part of the training observations, one speaks of an *overtrained* BDT. The BDT in such a case is not able to generalize the features, which it learned based on the training observations, to new observations.

To illustrate this, imagine a student preparing for an exam. The student learns the material covered in the lecture and uses old exams from the past years. The ideal case would be that the student learns the concepts of the material and is able to show this by solving similar exercises in the final exam. A student which corresponds to an overtrained BDT, would simply learn the answers of the previous exams by heart. This student is much more adapted to the training input and is able to solve previous exams without any mistake, but this will not help in the actual exam.

The goal of a successful boosted decision tree parameter optimization is therefore to find boosted decision tree parameters suited for the training observations at hand, such that the boosted decision tree performs best on new unseen observations.

Table 6.1: Setup of the parameter scan to illustrate the difference between training and test set.

Parameter	Value(s)
Input Variables	full set (47 variables)
Category	regular VBF
Background (training)	all
Background (test)	all
Signal (training)	VBF only
Signal (test)	VBF and ggF
Number of Trees	1, 2, 3, 6, 10, 20, 30, 60, 100, 200, 300, 600, 1000
Max Depth	10
Min Node Size	1 %

6.1.4 Boosted Decision Tree Parameter Scans

As discussed in the previous section, boosted decision trees tend to overtraining. It is therefore beneficial, to introduce two disjoint (statistically independent) sets of Monte Carlo events. One set is used to train the BDTs (hence *training set*) and the other set can be used to evaluate the performance (commonly named *test set*). By doing this, an overtrained BDT will (by the definition of overtraining) perform badly on the unseen MC set used for evaluation.

In this thesis the technical procedure of splitting the Monte Carlo events into two sets is performed based on the event number. One way to split the events is to define one Monte Carlo set that contains only events with an even event number, whereas the other set contains only events with odd event numbers.

As discussed in Sec. 6.1.2, boosted decision trees have several parameters to tune and control their training. The purpose of a *parameter scan* is to determine the best values of these parameters, such that the boosted decision tree is not overtrained but still as powerful as possible. In this context the best values are defined by maximizing the BDT performance on the test set according to some figure of merit. In this thesis the figure of merit of BDT performance has been chosen to be the significance determined by a likelihood fit with a limited set of systematic uncertainties as described in Sec. 6.2.

An easy way to perform a parameter scan is to choose a collection of different parameter configurations and train one boosted decision tree for each configuration. The BDTs can then be assessed by calculating the desired figure of merit. In this thesis a grid of parameter configurations is used in the parameter scans.

A BDT, selected based on its performance on the test set, can be assumed to be safe from overtraining. The principle, that the optimization (here the training) and the assessment of its outcome is done on two different Monte Carlo sets, is generally applicable, and will be re-encountered for other optimizations in this thesis as shown Sec. 6.1.5.

The purpose of the first study presented here, is to show a parameter scan and to illustrate the phenomena of overtraining at the same time. To perform this study

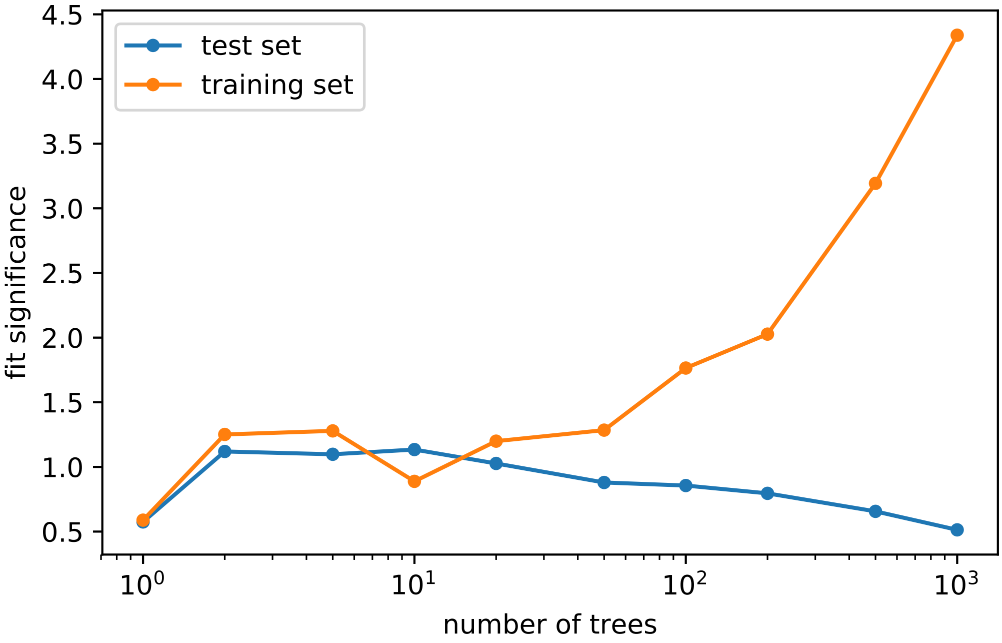

Figure 6.2: Comparison of the BDT performance when used on the training set and when used on an independent test set as a function of the number of trees in the BDT.

a parameter scan has been carried out. For this, only the parameter *number of trees* is varied. All other parameters are kept constant. Since a dedicated study to optimize the BDT input variables has not yet been described, this study uses all variables listed in Sec. 6.3. As stated in [30], boosted decision trees are robust against additional input variables. The setup of the parameter scan is summarized in Tab. 6.1. This study consists of a comparison of BDTs used on their training and test set. BDTs with different complexities are used, to see the dependence on the complexity of the classifier. The single parameter *number of trees* is chosen to represent and control the complexity of BDTs. A large number of trees leads to more boosting iterations and therefore a more complex BDT, which is more likely to experience overtraining.

The result of this study is illustrated in Fig. 6.2. The figure shows the expected behavior on the training and test set. For low BDT complexity the difference in performance between training and test set is small. Both increase until a certain complexity is reached. Beyond this point the BDT performance on the training set still increases, as the increasing complexity of the boosted decision tree adapts to small features in the training set. When this happens the boosted decision tree is overtrained and its performance decreases on the independent test set, since this does not feature the same fluctuations as the training set.

It should be noted that all points in the plot are subject to statistical fluctuations. This can be seen for ten trees, where the BDT used on the test set outperforms the same BDT on the training set. The points would scatter around a mean value, when the experiment was repeated on different training and test sets. Since the BDT training and the fit are both complicated and non-intuitive procedures and since multiple training and test sets are not available, it is difficult to estimate an uncertainty for the points.

6.1.5 Training, Validation and Test Sets

The procedure to test different configurations and select the best BDT based on its performance on the test set as outlined in Sec. 6.1.4, is only reliable when the number of parameter configurations is small. In the course of this thesis of the order of 10^5 boosted decision trees are trained. As the number of parameter configurations increases the parameter scan itself becomes an extensive optimization which can feature its own overtraining behavior.

A boosted decision tree, which is selected because it is the best BDT in a parameter scan, can bias the analysis. The selected BDT might be able to separate the background and signal particularly well by chance, only when used on this particular test set. In fact, it should not be surprising that such a bias can be introduced, by selecting the *best* boosted decision tree on the Monte Carlo set, on which it will be used in the analysis. If the selected BDT exhibits such a bias, the boosted decision tree will perform better on Monte Carlo compared to data and therefore will lead to discrepancies in the output distributions of the boosted decision tree.

The usual approach in the machine learning community to mitigate this problem is to introduce a third disjoint (statistically independent) Monte Carlo set. Since the Monte Carlo statistics is limited and a new third set can not be added easily, this three-fold instead of two-fold splitting has to be done before any training or grid scanning is performed. The three sets are used as follows:

- The *training set* has the same function as in the previous section to train all BDTs in the parameter scan.
- The second set is commonly named *validation set*. This set takes the role of the test set in the approach from the previous section. That is, all the trained BDTs are used on this set to assess their performance. Based on this evaluation, the best performing BDT is selected.
- Once a boosted decision tree has been chosen, the *test set* can be used (not to be confused with the test set in the previous section). All the events in the test set are then classified with the selected boosted decision tree. The same classification is carried out with measured data. The resulting distributions can be used in the final fit of the analysis.

By introducing two separate sets for the parameter optimization and the final fit, the potential bias towards better performances on the optimization set can not affect the agreement of data and Monte Carlo in the analysis. Neither the Monte Carlo events in test set nor the data events have been used to train or optimize the BDT. In a way, this procedure is similar to *blinding* of interesting parts of data until the analysis is fixed. The difference, however, is that here parts of the Monte Carlo (the test set) and not data is blinded until the analysis is fixed.

The scheme proposed in this section overcomes the issue of a bias introduced by the parameter scan itself. However, the scheme introduces a new complication. In general the training statistics is crucial in multivariate analyses. By splitting the Monte Carlo set into three disjoint, equally large sets, the training statistics is reduced by 33% compared to the simple approach of only two sets. The situation can be improved by employing k-fold cross validation.

6.1.6 k-fold Cross Validation

In Run 1 the Monte Carlo events were split into two sets of MC events, based on the oddness of the event number. A boosted decision tree was trained on the even Monte Carlo set and was then used on the odd event set. Another BDT was trained on the odd set and used on the even set. By doing this the BDTs could be trained on 50% of all MC events. In the analysis, however, 100% of all Monte Carlo events can be used. Each event is classified by the BDT which was not using this event during training. This procedure can be referred to as 2-fold cross validation, because it involves two independently trained BDTs. In the following analysis this method is generalized to k-fold cross validation with arbitrary k and modified to incorporate the principle of disjoint training, validation and test sets.

To use k-fold cross validation, the MC events are split into k disjoint and statistically independent sets $s = 0, \ldots, (k-1)$. Similarly to the previous technical procedure, for this thesis the splitting is done based on the event number n. The index i of the set to which an event with event number n belongs is determined by

$$i = n \mod k. \tag{6.9}$$

No correlation between event number and any kinematic variable could be found. The Monte Carlo sets have been checked to be equivalent given their statistical uncertainty.[2]

For each parameter configuration in a parameter scan k boosted decision trees are trained. The total number of boosted decision trees to train in a parameter scan is therefore multiplied by k. Given a parameter configuration, all k BDTs are trained on different combinations of MC sets. BDT_i uses all Monte Carlo sets except set $s = i$ and $s = (i+1) \mod k$. Consider for example $k = 10$. For each parameter configuration ten BDTs are trained. The first one, with index $i = 0$, uses the sets $s = 2, \ldots, 9$ for its training. This means 80% of the Monte Carlo statistics are available during training. The Monte Carlo sets $s = 0$ and $s = 1$ have not been seen by $\mathrm{BDT}_{i=0}$. The association between Monte Carlo sets and boosted decision trees is illustrated in Fig. 6.3.

One of the two remaining Monte Carlo sets can be used to evaluate the boosted decision tree and therefore function as validation set, whereas the last remaining set takes the role of the test set and is used to produce the final BDT score distributions for the analysis. While using the boosted decision trees on the validation sets or test sets, the k boosted decision trees can be thought of as a single compound classifier. The combined classifier is used on the full Monte Carlo set. The event number of each event will decide which internal boosted decision tree is used to classify the event. This is the same as adding the k BDT score distributions. In the following

[2] From a physical perspective such a correlation seems impossible. However, during the preparation of the Monte Carlo samples, it can not be excluded that the event number has already been used and that events are treated differently based on the event number. The ideal case would be to perform the splitting using a variable from a independent random generator. This is however not possible in an easy way while maintaining reproducibility and cross framework compatibility.

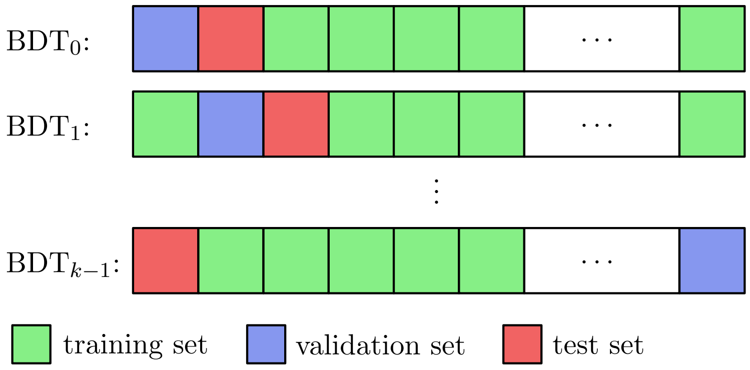

Figure 6.3: Illustration of the association between Monte Carlo set and boosted decision tree in k-fold cross validation.

the term boosted decision tree is also used to refer to this compound classifier. It is apparent from the context when a single boosted decision tree inside the compound boosted decision tree is meant.

6.2 Assessment of Boosted Decision Trees

The sensitivity of an analysis can depend to a large extent on systematic uncertainties. Selecting the most promising boosted decision tree based on a figure-of-merit which takes only statistical uncertainties into account, does not guarantee that the performance of the BDT is also good once systematic uncertainties are introduced.

To get a more reliable estimation of the BDT performance during the parameter optimization, a likelihood fit of the boosted decision tree output score distribution is performed. Only a limited set of systematic uncertainties is considered in this fit, since calculating the BDT score distributions for systematic variations is a (CPU) time consuming process. The list of systematic uncertainties has been created by considering the ranking of systematic variations by their impact on the coupling strength in the cut-based analysis. Additionally, systematic variations, that are expected to be important from a physics point-of-view, are also included. The variations included in the fit are listed in Tab. 6.2. Descriptions of the systematic uncertainties and more information about the fit procedure can be found in Chap. 7.

This reduced version of the fit does not use any control regions to constrain the background processes. The parameter optimization is carried out separately for the two categories VBF and Boosted. The fit to assess the boosted decision tree during the optimization is therefore also performed independently for the two categories.

6.3 Input Variables

The selection of input variables of the boosted decision tree is an important step in the optimization of a BDT-based multivariate analysis. Unlike other machine learning algorithms such as (deep) neutral networks, a boosted decision tree is not able to learn how to derive new variables [28, 29]. It is therefore beneficial to preprocess the input variables and combine them to derive variables which capture

Table 6.2: List of systematic variations used in the fit to assess the boosted decision tree performance. The second column indicates, whether the systematic variation consists of a up and down variation.

Name of the Systematic Variation	Type
`btag_light_0`	up/down
`JER`	
`JES_EffectiveNP_1`	up/down
`JES_EffectiveNP_2`	up/down
`JES_EffectiveNP_3`	up/down
`MET_SoftTrk_ResoPara`	
`MET_SoftTrk_ResoPerp`	
`PU_PRW`	up/down
`TAU_EFF_ELEOLR_TRUEELECTRON`	up/down
`TAU_TES_INSITU`	up/down

the event topologies. The definition of derived quantities has been guided by the results from Run 1 [15]. Additionally, other variables such as the hadron collider moments have been added [27]. The derived quantities have been introduced in Sec. 5.2. The following list summarizes all the variables that have been considered as input variables of the BDT.

- Hadron collider moments: $h_1, \ldots, h_8$.
- Jet(s): η^{j_i}, ϕ^{j_i}, $p_{\mathrm{T}}^{j_i}$ for $i = 0, 1$, $\eta^{j_0} \cdot \eta^{j_1}$, $\Delta\eta^{jj}, \sum p_{\mathrm{T}}^{j}, m^{jj}$
- Di-tau and Higgs boson system: $x_{0,1}^{\mathrm{collin}}$, $\Delta\eta^{\mathrm{lep\,had}}$, $\Delta\phi^{\mathrm{lep\,had}}$, $\Delta p_{\mathrm{T}}^{\mathrm{lep\,had}}$, $\Delta R^{\mathrm{lep\,had}}$, m_{MMC}, m_{T}, p_{T}^{H}, $p_{\mathrm{T}}^{lep}/p_{T}^{\mathrm{had}}$, $\sum p_{\mathrm{T}}$, $p_{\mathrm{T}}^{\mathrm{total}}$, m_{vis}
- Lepton: η^{lep}, ϕ^{lep}, $p_{\mathrm{T}}^{\mathrm{lep}}$, $\ell\,\eta$ centrality and the tight lepton identification criteria
- Missing transverse energy: $E_{\mathrm{T}}^{\mathrm{miss}}$, $E_{\mathrm{T}}^{\mathrm{miss}}\phi$ centrality and ϕ of $E_{\mathrm{T}}^{\mathrm{miss}}$,
- Number of electrons, muons and jets
- Tau: η^{had}, ϕ^{had}, $p_{\mathrm{T}}^{\mathrm{had}}$

As discussed in Sec. 6.1.1, each node in the trees is constructed by examining the full variable list and choosing the variable and cut position which maximizes the separation gain. It is therefore possible to train boosted decision trees and inspect their structure to see which variables are important.

6.3.1 Variable Selection

The study presented in Sec. 6.1.4 used the full variable set shown in Sec. 6.3. Although BDTs are robust and ignore variables which do not provide additional information, it is, however, advisable to limit the input variables to a minimal set to

Table 6.3: Setup of the parameter scan to determine the optimal set of input variables. The same setup is used for VBF and Boosted.

Parameter	Value(s)
Cross Validation Sets k	10
Input Variables	full set (47 variables)
Training Region	regular VBF / Boosted region
Background (training)	all
Background (validation)	all
Signal (training)	VBF / ggF for VBF / Boosted
Signal (validation)	VBF / ggF for VBF / Boosted
Number of Trees	10, 20, 50, 100, 200, 400, 800, 1000
Max Depth	10
Shrinkage	0.03, 0.1, 0.3, 0.8
Min Node Size	0.1, 0.2, 0.5, 1, 2, 5, 10, 20 %

reduce computation time and to limit the effort it takes to validate the input modeling, since this includes the correlations between the input variables (*i.e.* $\mathcal{O}(n^2)$ correlation pairs).

A sensible set of input variables is determined with the use of a BDT parameter scan as outlined in Tab. 6.3. BDTs with various parameter configurations are trained and assessed with a fit on the validation set as has been described in Sec. 6.2. The software library TMVA offers a measure to assess the importance of an input variable for a given BDT. As defined in [30], the *importance* of an input variables is derived from the sum of squared separation gain that the variable achieves at a node in the BDT weighted by the number of events that are affected by the node. Separation gain is quantified by the gini index [30]. A global variable rankings are derived from the parameter scan by averaging the importance values for each variable over all BDTs. The parameter scan is performed separately for VBF and Boosted, as it is assumed that different sets of input variables provide an optimal input set for the two categories due to the different event topology. The resulting full variable ranking is listed in Tables B.1 and B.2.

The top ranked variables are similar to the input variable set used in Run 1 [15]. The variable usage in Run 1 is listed in Tab. 6.4. Several variables in the top of the ranking are strongly correlated to other highly ranked variables, such as $\Delta R^{\mathrm{lep\,had}}$, $\Delta\eta^{\mathrm{lep\,had}}$ and $\Delta\phi^{\mathrm{lep\,had}}$. It is assumed that the effect on the BDT performance is only marginal, if these variables are collapsed into a single variable. If this procedure is repeated, the top of the ranking consists almost exclusively of the variable set used in Run 1, which has been optimized as described in Ref. [15]. Since the variable ranking indicates a similar choice, the analysis presented in this thesis uses the variable set from Run 1.

Table 6.4: Input variable set used in Run 1 and in this thesis. The dots indicate the usage of variable in a category.

Variable	VBF	Boosted
m_{MMC}	●	●
ΔR^{lephad}	●	●
m_{T}	●	●
$E_{\text{T}}^{\text{miss}} \phi$ centrality	●	●
m^{jj}	●	
$\Delta\eta^{jj}$	●	
$\ell\, \eta$ centrality	●	
$\eta^{j_0} \cdot \eta^{j_1}$	●	
$p_{\text{T}}^{\text{total}}$	●	
$p_{\text{T}}^{\text{lep}}/p_{T}^{\text{had}}$		●
$\sum p_{\text{T}}$		●

6.3.2 Modeling of Input Variables

Figures 6.4 and 6.5 show the distributions of input variables in the VBF and Boosted categories, respectively. The distributions show good agreement between data and the background estimation. Since a boosted decision tree is capable of exploiting correlations between the input variables, the agreement of the correlations between data and Monte Carlo have also been checked. No significant discrepancy could be found. This set of input variables can therefore be used to train the boosted decision tree used for this multivariate analysis.

Figures 6.6 and 6.7 show the linear correlation coefficients between the input variables for the VBF and Boosted categories respectively. The boosted decision trees can utilize differences in the correlations between signal and background to achieve a higher separation between signal background compared to the cut-based analysis.

6.4 Generator Filter and Training Statistics

An essential ingredient for successful machine learning is to have a large set of training observations. In this analysis this means to have a large set of Monte Carlo events, which can be used to train the boosted decision trees. From earlier studies it is known that the statistics of the signal processes is a limiting factor. There are various techniques to mitigate this problem.

The most obvious method is to enhance the training statistics by generating more Monte Carlo events. Since this is a (computationally) time consuming task, the event generation should be tailored to this analysis. This can be done by applying selections at an early stage of the generation process. This allows to reject events, if they will not be selected in the analysis, and thus this saves the computation time of all subsequent steps for these events. This early selection is generally referred to

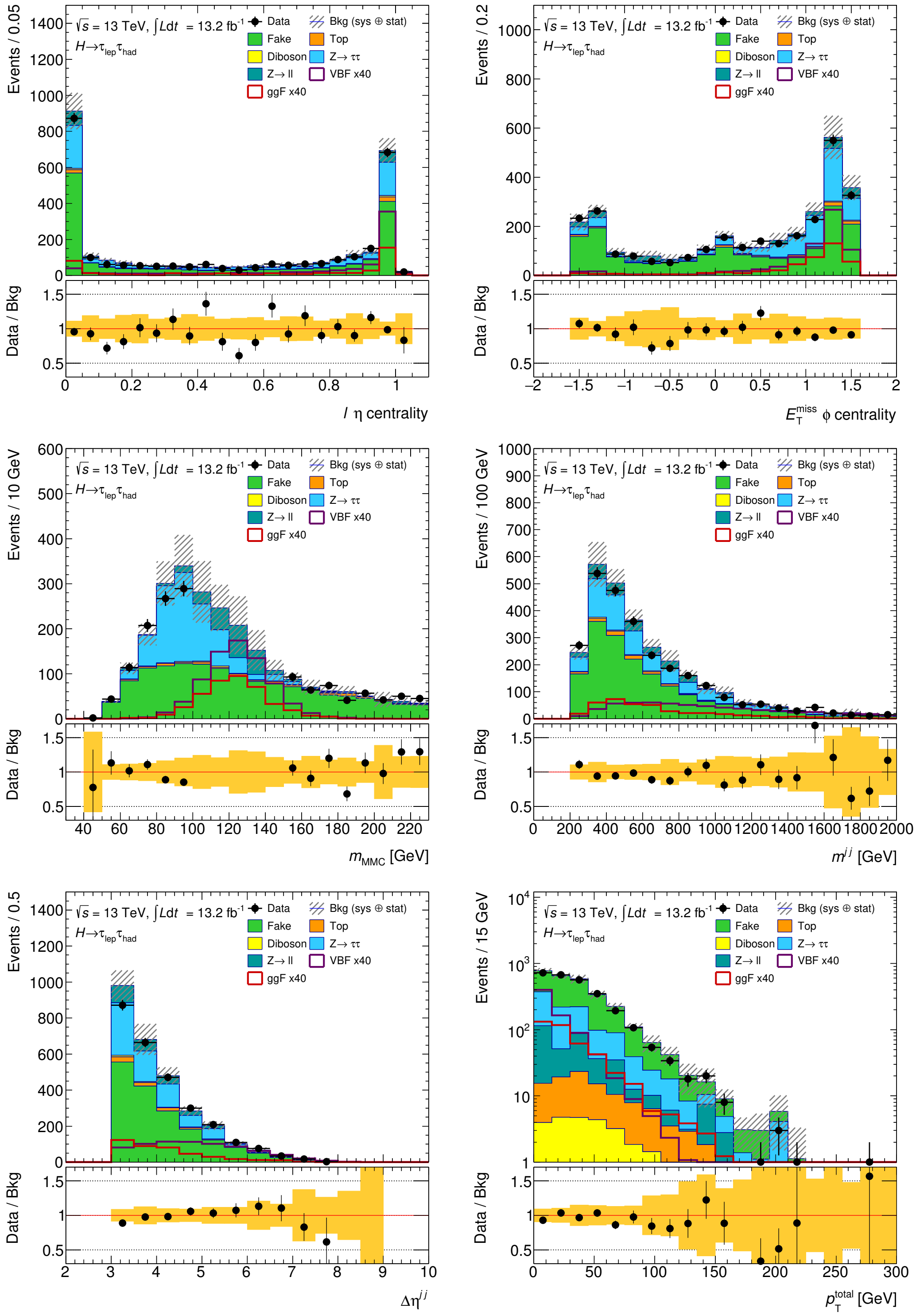

Figure 6.4: Selected input variables of the VBF boosted decision tree. The top row shows $\ell\,\eta$ centrality (left) and $E_{\mathrm{T}}^{\mathrm{miss}}\phi$ centrality (right). The middle row shows m_{MMC} with blinded data in the signal-sensitive region $100\,\mathrm{GeV} < m_{\mathrm{MMC}} < 150\,\mathrm{GeV}$ (left) and the invariant mass of the jets m^{jj} (right). The bottom row shows $\Delta\eta^{jj}$ (left) and $p_{\mathrm{T}}^{\mathrm{total}}$ (right).

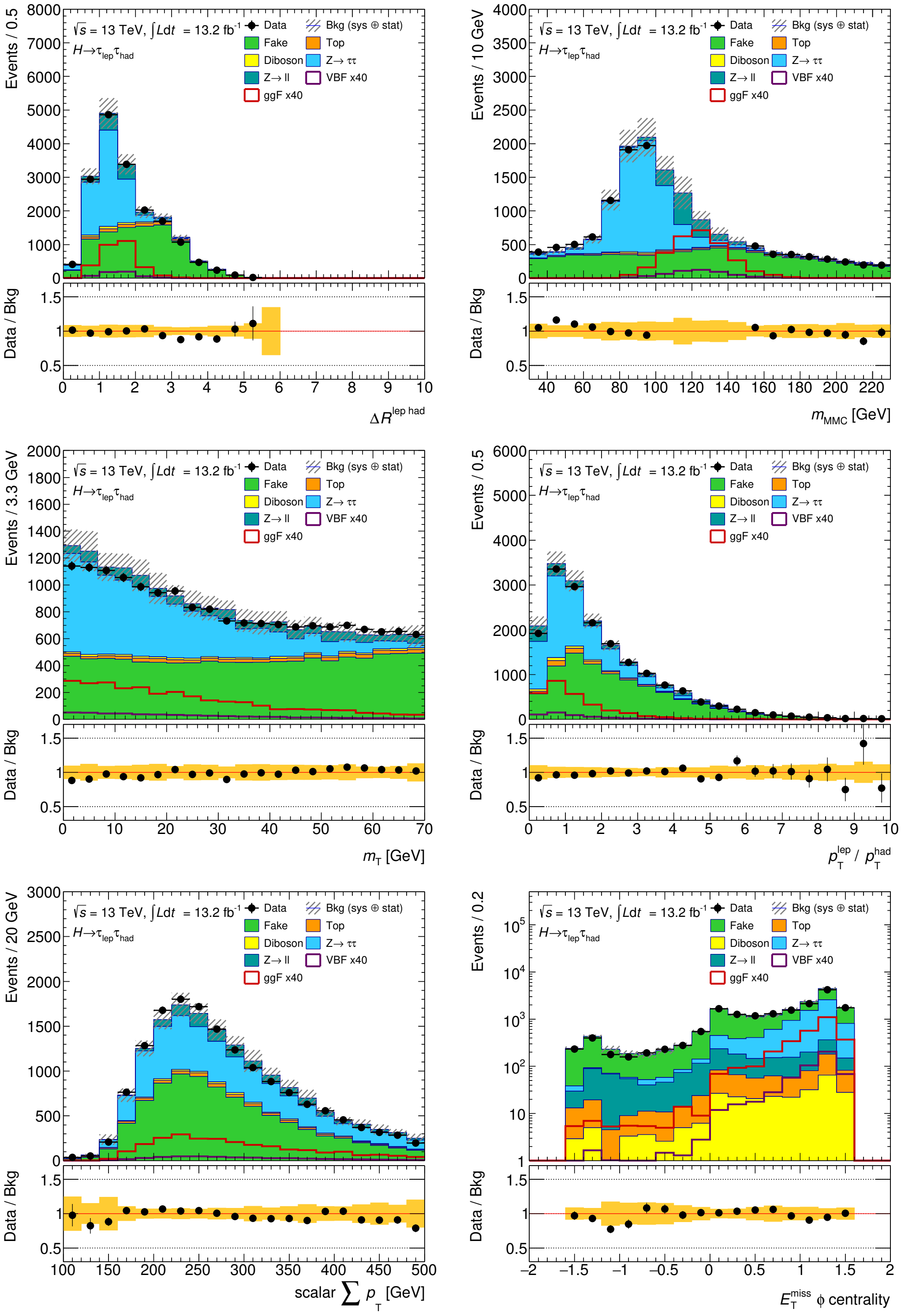

Figure 6.5: All input variables of the Boosted boosted decision tree. The top row shows $\Delta R^{\text{lep had}}$ (left) and m_{MMC} with blinded data in the signal-sensitive region $100\,\text{GeV} < m_{\text{MMC}} < 150\,\text{GeV}$ (right). The middle row shows the transverse mass m_{T} (left) and the ratio of the transverse momenta of the tau and the lepton $p_{\text{T}}^{\text{lep}}/p_{T}^{\text{had}}$ (right). The bottom row shows the scalar sum of the transverse momenta $\sum p_{\text{T}}$ (left) and $E_{\text{T}}^{\text{miss}}\,\phi\,$centrality (right).

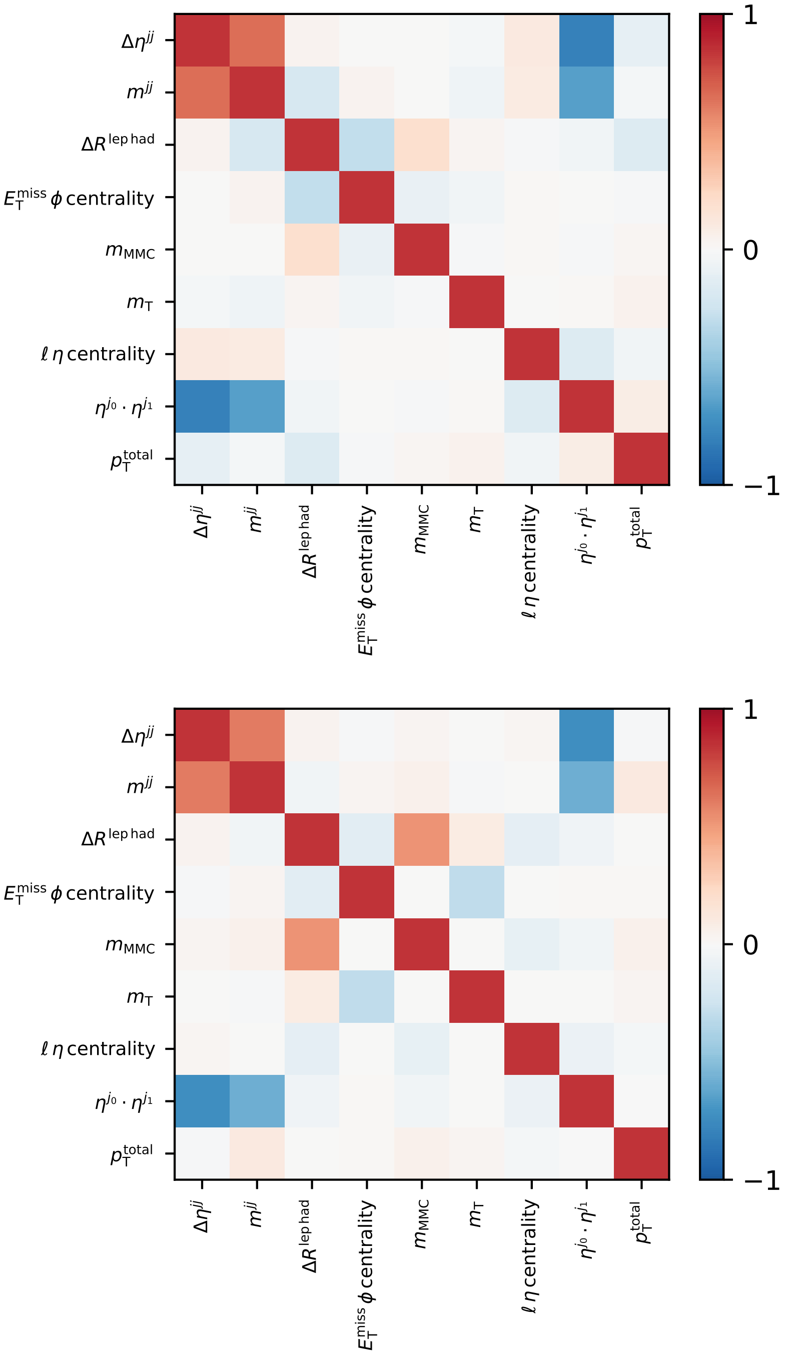

Figure 6.6: Visualization of linear correlations coefficients between the input variables in the VBF category for the signal class (top) and the background class (bottom).

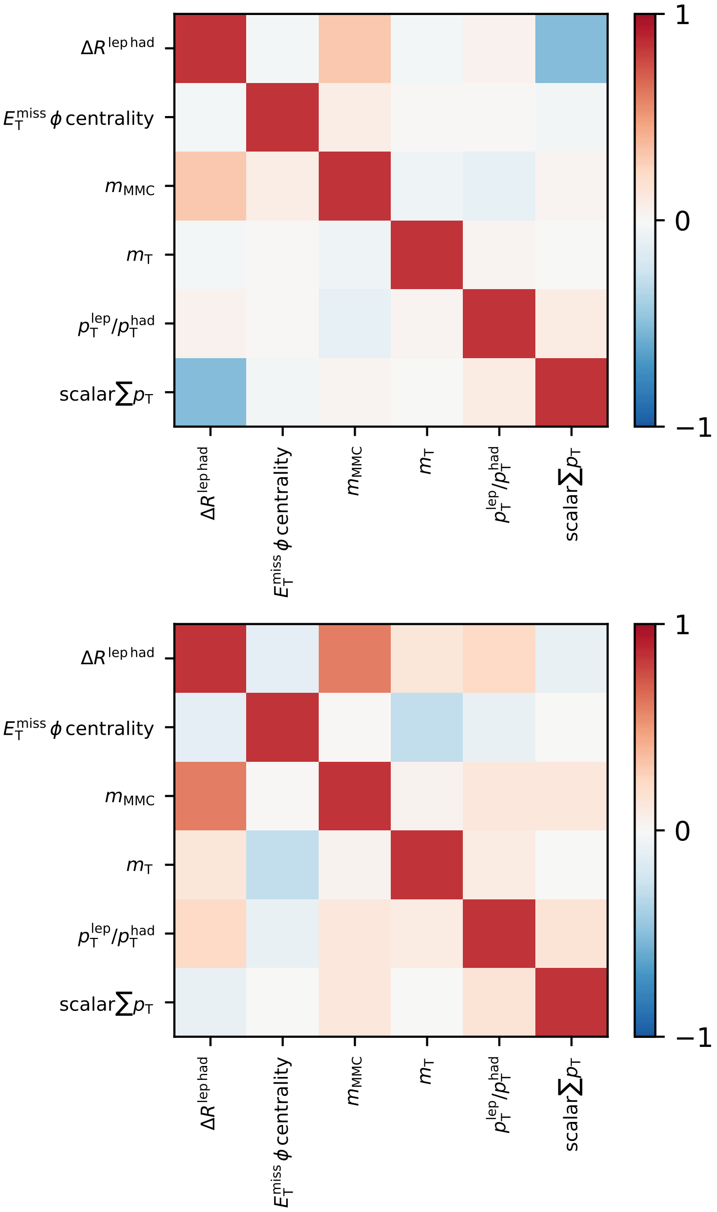

Figure 6.7: Visualization of linear correlations coefficients between the input variables in the Boosted category for the signal class (top) and the background class (bottom).

as a *Generator Filter* and will be described in more detail in this section. Since the filtered Monte Carlo samples are produced privately and have not been approved, the filtered Monte Carlo events will only be used as training input. The events from the filtered Monte Carlo production are removed when the Monte Carlo set is used as a validation or test set.

The outline of the individual steps and stages required for Monte Carlo event generation has been described in Sec. 4.4. Section 6.4.1 describes the generator filter developed for this application and Sec. 6.4.2 compares the filtered signal sample to the officially produced inclusive signal samples. Finally, Section. 6.4.3 studies the effect of an increase in training statistics

6.4.1 Development of a Generator Filter

A generator filter is implemented at the step of event generation. After the simulation of the initial hard scattering process, the generator filter decides whether the subsequent detector response should be simulated. If an event is discarded, because it will not pass the event selection of the analysis, the detector simulation and reconstruction do not have to be carried out.

The filter developed for this thesis is built upon the configuration of the official sample production. However, in contrast the official samples, ATLFAST-II is used to simulate the detector response. The official samples are filtered by selecting the final state $\tau_{\text{lep}}\tau_{\text{had}}$ of each $H \to \tau\tau$ decay. The filter selections described in the following are added to this already existing configuration.

The goal of this generator filter is to enhance the Boosted signal region with gluon fusion signal samples. The Boosted category has the signature of high transverse momentum of the Higgs boson, which can be used to identify event candidates at an early stage. A generator filter which enhances this signal statistics in the VBF category is also desirable and attainable, but has not been investigated in the course of this thesis.

A complication arises because an event filter at generator level only has access to the truth information of the particles. In the analysis, however, the event selection is based on reconstructed quantities. Optimally the truth information and the reconstructed information is identical. The reconstructed quantities differ from the truth information, due to detector effects, especially energy resolution and energy scale, but also due to misidentification. To distinguish the truth quantities from the reconstructed quantities, the symbol for truth quantities is annotated by a hat, for example $\hat{p}_{\text{T}}^{a}$ denotes the truth transverse momentum of particle a.

One approach to mitigate this difficulty is to check the distributions of the truth quantities after applying the regular analysis selections, which are based on the reconstructed quantities. By studying the truth distributions one can select a cut value for the truth quantities, which will not affect the events in the analysis categories. This approach has been taken to implement cuts on the truth transverse momentum of the tau $\hat{p}_{\text{T}}^{\text{had}}$ and the lepton $\hat{p}_{\text{T}}^{\text{lep}}$. Based on the distributions in Fig. 6.8, a threshold of 20 GeV has been introduced for both quantities.

Another approach has been taken to implement a generator filter cut for the transverse momentum p_{T}^{H} of the Higgs boson. The truth $\hat{p}_{\text{T}}^{H}$ distribution at generator level is a steeply falling curve similarly to the $\hat{p}_{\text{T}}^{H}$ distribution at preselection level shown in Fig. 6.9. The offline selection for transverse momenta of the Higgs boson

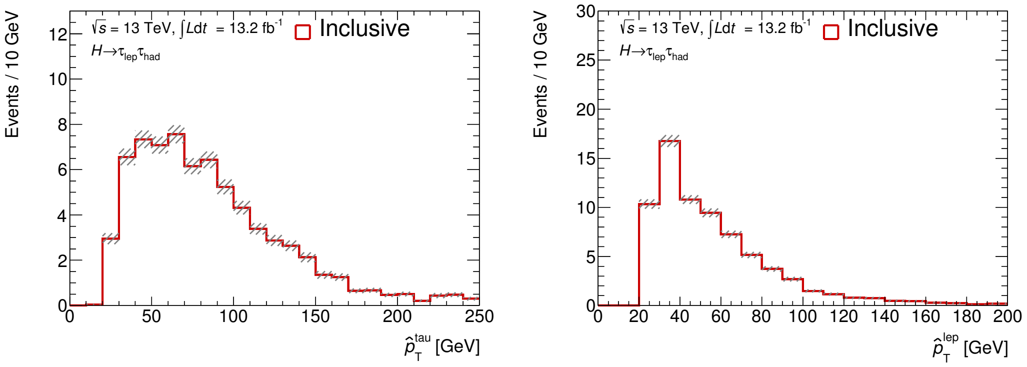

Figure 6.8: Truth transverse momentum of the tau $\hat{p}_{\mathrm{T}}^{\mathrm{had}}$ (left) and the lepton $\hat{p}_{\mathrm{T}}^{\mathrm{lep}}$ (right) after applying the Boosted selection criteria of the multivariate analysis. The red line shows the official inclusive ggF production.

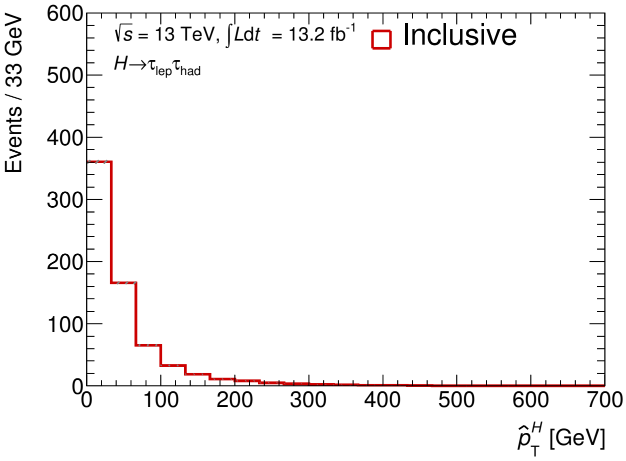

Figure 6.9: Truth transverse momentum of the Higgs $\hat{p}_{\mathrm{T}}^{H}$ boson at preselection level. The red line shows the official inclusive ggF production.

Table 6.5: Summary of all additional cuts. This includes the generator filter cuts itself, the cuts necessary to combine the filtered Monte Carlo events with inclusive Monte Carlo events, and the definition of the new control region.

Production	SR Cut	CR Cut	Applied at
ggF Filtered	$\hat{p}_{\mathrm{T}}^{\mathrm{had}} > 20\,\mathrm{GeV}$	$\hat{p}_{\mathrm{T}}^{\mathrm{had}} > 20\,\mathrm{GeV}$	generator level
	$\hat{p}_{\mathrm{T}}^{\mathrm{lep}} > 20\,\mathrm{GeV}$	$\hat{p}_{\mathrm{T}}^{\mathrm{lep}} > 20\,\mathrm{GeV}$	generator level
	$\hat{p}_{\mathrm{T}}^{H} > 100\,\mathrm{GeV}$	$\hat{p}_{\mathrm{T}}^{H} > 100\,\mathrm{GeV}$	generator level
ggF Inclusive	$\hat{p}_{\mathrm{T}}^{H} < 100\,\mathrm{GeV}$	$\hat{p}_{\mathrm{T}}^{H} > 100\,\mathrm{GeV}$	analysis level

Table 6.6: Comparison of the raw event count in the boosted analysis category for gluon fusion signal samples. The numbers are also listed separately for two different parts of phase space separated by the truth $\hat{p}_{\mathrm{T}}^{H}$.

Sample	Boosted	with $\hat{p}_{\mathrm{T}}^{H} > 100\,\mathrm{GeV}$	with $\hat{p}_{\mathrm{T}}^{H} < 100\,\mathrm{GeV}$
ggF Inclusive	6534	5869	665
ggF Filtered	63502	63502	0

is $p_{\mathrm{T}}^{H} > 100\,\mathrm{GeV}$. The edge at 100 GeV in the distribution of the truth $\hat{p}_{\mathrm{T}}^{H}$ is smeared out due to detector effects. As indicated by the red line in Fig. 6.10 a cut at generator level has to be at about 40 GeV in order to avoid biasing the signal regions. According to Fig. 6.9 a truth cut at 40 GeV will not suppress unattained events as much as desired.

To profit from the high suppression rate of a truth-level selection of $\hat{p}_{\mathrm{T}}^{H} > 100\,\mathrm{GeV}$ and to prevent biasing the signal region at the same time, the filtered Monte Carlo sample is combined with the official inclusive Monte Carlo sample. To avoid an excess of events with truth $\hat{p}_{\mathrm{T}}^{H} > 100\,\mathrm{GeV}$ by "double counting" them, the phase space covered by the filtered Monte Carlo samples is removed from the official production by introducing a truth cut $\hat{p}_{\mathrm{T}}^{H} < 100\,\mathrm{GeV}$ exclusively for the official production. The additional filter cuts are summarized in Tab. 6.5.

The filter efficiency has been determined to $\epsilon = 0.02061$. The filter efficiency of $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ final state filter, which is used for the official production, is $\epsilon = 0.4548$. Thus, for the same number of produced events, the filtered MC sample provides a statistical power a factor of 22 larger compared to the official sample. In total 1.8 million filtered events have been produced. About 40 million inclusive events had to be generated to achieve the same statistical power. The number of events in the signal region of the inclusive and filtered productions are listed in Tab. 6.6. The filtered MC sample achieves a ten-fold increase of events in the signal region.

6.4.2 Sample Validation

To validate the privately produced filtered Monte Carlo samples, the modeling is compared to the official inclusive production. Firstly, the border in phase space, where inclusive and filtered samples are stitched together, is analyzed, to check if the transition point is smooth. Secondly, a new control region is defined for this

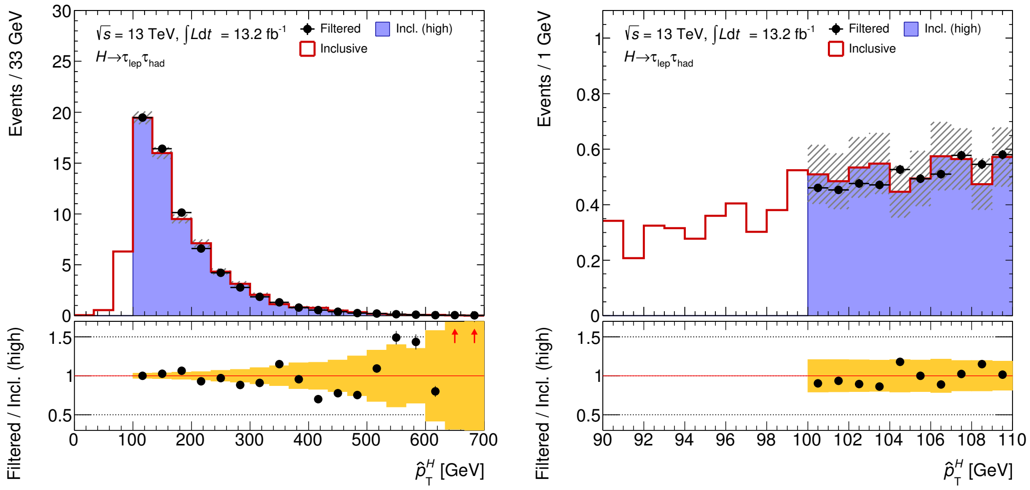

Figure 6.10: The two distributions show the truth transverse momentum of the Higgs $\hat{p}_{\mathrm{T}}^{H}$ in the Boosted category of the multivariate analysis. The two plots differ only by their x-axis range. The red line corresponds to the inclusive ggF production. The blue area corresponds to the official production with $\hat{p}_{\mathrm{T}}^{H} > 100\,\mathrm{GeV}$. The error bands corresponds to the statistical uncertainty of the inclusive sample in the control region. The black dots represent the filtered production. The ratio plot shows the yield ratio of filtered production (black dots) over the official production with $\hat{p}_{\mathrm{T}}^{H} > 100\,\mathrm{GeV}$ (blue).

study, by inverting the truth cut to $\hat{p}_{\mathrm{T}}^{H} > 100\,\mathrm{GeV}$. In this special CR it is possible to compare the modeling of the filtered Monte Carlo to the official Monte Carlo production. The cut applied for the control region is also listed in Tab. 6.5.

Figure 6.10 shows the transition point at $\hat{p}_{\mathrm{T}}^{H} = 100\,\mathrm{GeV}$ between the inclusive Monte Carlo and filtered Monte Carlo in a $\hat{p}_{\mathrm{T}}^{H}$ histogram. Below 100 GeV the official MC will be used in the analysis, above 100 GeV the filtered Monte Carlo will be used. The blue area corresponds to events from the inclusive sample in the control region with $\hat{p}_{\mathrm{T}}^{H} > 100\,\mathrm{GeV}$. Judging from Fig. 6.10, the transition across the border in phase space is smooth. No discontinuity or kink can be found at $\hat{p}_{\mathrm{T}}^{H} = 100\,\mathrm{GeV}$, which is not attributable to the statistical uncertainties. The attention should also be focused on the size of the error bands of the inclusive sample compared to the error bar of the filtered sample. The smaller error bars, which are to small to be visible for most points, in the region of filtered samples, indicate the increase in statistics, which is gained by using the filtered samples.

In Fig. 6.11 the modeling of the mass variables in the filtered samples is compared to the modeling in the official sample. A clear slope is visible in the ratio plots for m_{vis} and m_{MMC}. The modeling of other variables does not show any crucial mismodeling. The differences in the model between filtered and the official inclusive production is likely to stem from the fact, that the filtered sample has been simulated with ATLFAST-II, whereas the inclusive Monte Carlo was produced with Full Simulation.

To improve the modeling of the mass variables m_{vis} and m_{MMC} several reweighting schemes have been tested. In each scheme a kinematic variable v is chosen and the yield ratio of the inclusive sample over the filtered sample binned in this variable is fitted by a linear fit of the form $w(v) = \alpha + \beta v$. The weights of the filtered Monte

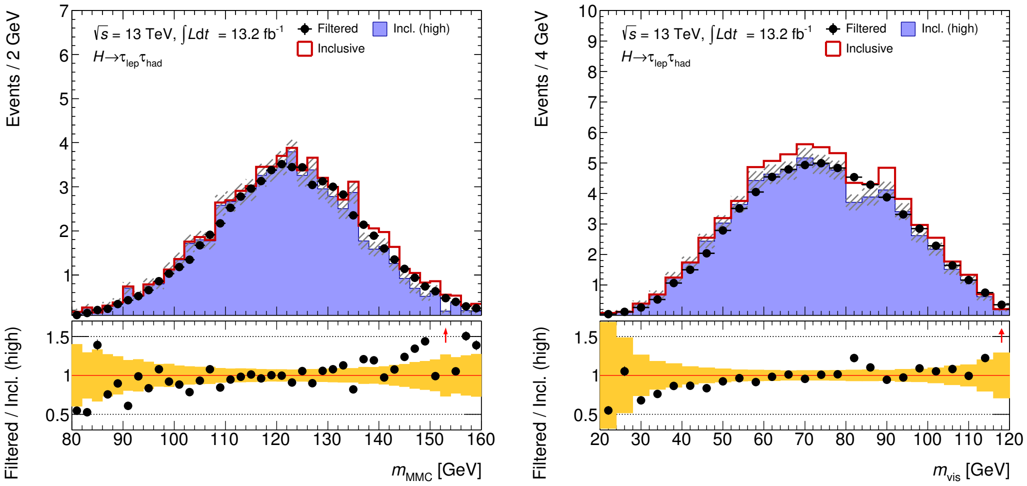

Figure 6.11: The mass m_{MMC} (left) and the visible mass m_{vis} in the Boosted category of the multivariate analysis. The red line corresponds to the inclusive ggF production. The blue area corresponds to the official production with $\hat{p}_{\text{T}}^{H} > 100\,\text{GeV}$. The error bands corresponds to the statistical uncertainty of the inclusive sample in the control region. The black dots represent the filtered production. The ratio plot shows the yield ratio of filtered production (black dots) over the official production with $\hat{p}_{\text{T}}^{H} > 100\,\text{GeV}$ (blue). The events are not reweighted.

Carlo events are then multiplied with $w(v)$. This procedure has been performed independently for $v = p_{\text{T}}^{\text{lep}}, p_{\text{T}}^{\text{had}}, \Delta R^{\text{lep had}}$ with no noticeable improvement for m_{vis} and m_{MMC}.

To improve the mass modeling it has been decided to use the above reweighting scheme to reweight m_{MMC} directly. The new event weights of the filtered MC sample are obtained by the multiplication of the events weights with

$$w(m_{\text{MMC}}) = 1.84028 - 0.00703632 \cdot \frac{m_{\text{MMC}}}{\text{GeV}}, \tag{6.10}$$

as described above. The reweighted mass distributions are shown in Fig. 6.12. The approach to correct m_{MMC} might seem dubious at first glance. However, it should be noted that this procedure does not involve any data events. The ATLFAST-II sample is merely adjusted to fit the sample produced with Full Simulation. Also one should consider that the filtered Monte Carlo will only be used for the BDT training. These events will not enter the analysis in any other way. No mismodeling in other variables has been found, which seemed to be responsible for a mismodeling in the derived quantity m_{MMC}. If there is a residual mismodeling after this correction procedure in other variables, which might have been the initial cause of the mass discrepancies, it will not result in physically wrong conclusions in the final analysis. In the worst case, a mismodeling in the training set can lead to a sub-optimal classifier, which then leads to a less sensitive analysis. The result of the analysis, however, will still be valid.

Figure 6.13 shows the modeling of important kinematic variables after applying the weights from Eq. 6.10. No major discrepancy in the relevant regions can be found. The reweighted filtered signal samples will therefore be used for the MVA training.

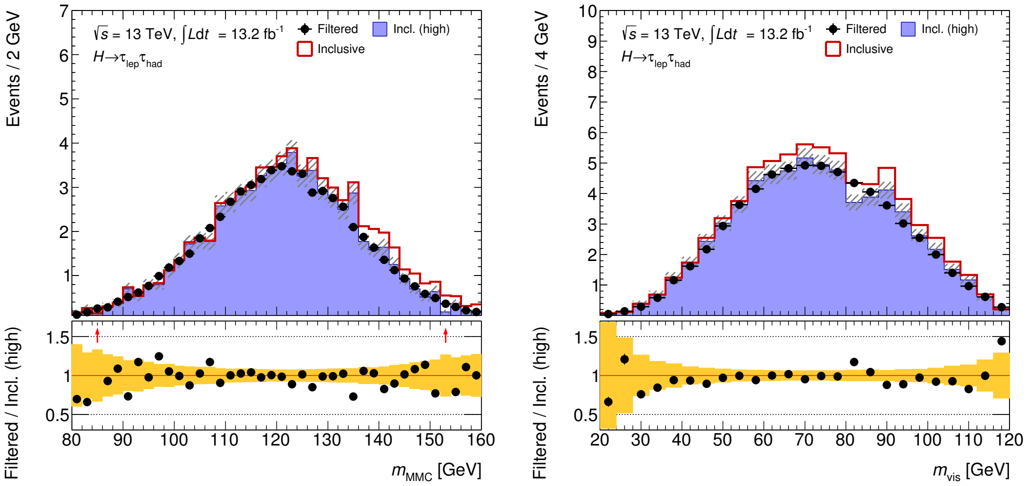

Figure 6.12: The mass m_{MMC} (left) and the transverse mass m_{T} in the Boosted category of the multivariate analysis. The red line corresponds to the inclusive ggF production. The blue area corresponds to the official production with $\hat{p}_{\mathrm{T}}^{H} > 100\,\mathrm{GeV}$. The error bands corresponds to the statistical uncertainty of the inclusive sample in the control region. The black dots represent the filtered production. The ratio plot shows the yield ratio of filtered production (black dots) over the official production with $\hat{p}_{\mathrm{T}}^{H} > 100\,\mathrm{GeV}$ (blue). The events of the filtered production are reweighted to match the m_{MMC} ratio plot.

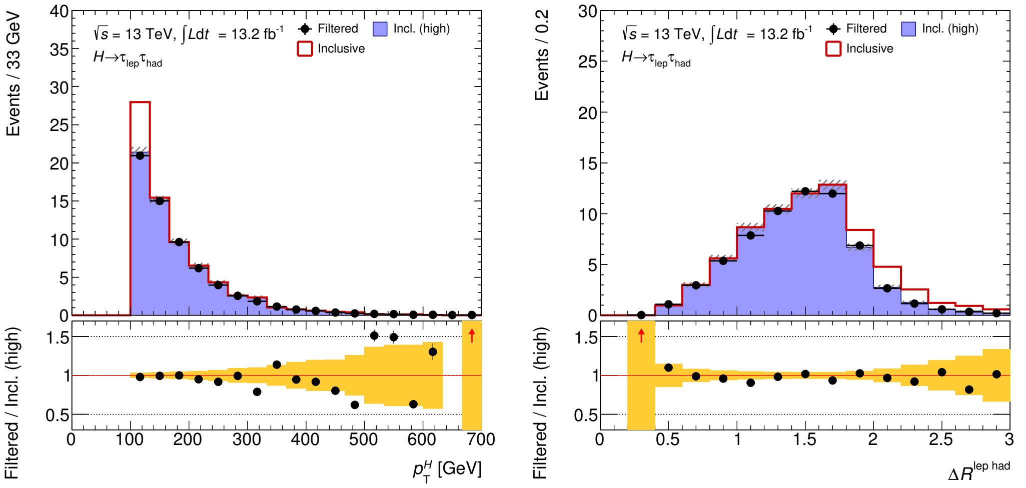

Figure 6.13: Important kinematic variables in the Boosted category in the MVA. The figure shows the reconstructed transverse momentum of the Higgs boson p_{T}^{H} (left) and the reconstructed angular difference between the lepton and the tau $\Delta R^{\mathrm{lep\,had}}$. The red line corresponds to the inclusive ggF production. The blue area corresponds to the official production with $\hat{p}_{\mathrm{T}}^{H} > 100\,\mathrm{GeV}$. The error bands corresponds to the statistical uncertainty of the inclusive sample in the control region. The black dots represent the filtered production. The ratio plot shows the yield ratio of filtered production (black dots) over the official production with $\hat{p}_{\mathrm{T}}^{H} > 100\,\mathrm{GeV}$ (blue). The events of the filtered production are reweighted to match the m_{MMC} ratio plot.

Table 6.7: Comparison of the raw events in Run 1 and Run 2. The second and third columns compare the number of events per physics process in each category for Run 2. The last two columns show the event counts from Run 1 as written in [15]. "–" is used if the information is not available. The numbers refers to raw, unweighted event counts. The table is limited to the most important processes. For Boosted in Run 1 the signal event count is only available as the sum of ggF and VBF. The process ggF denotes the official inclusive production.

	Run 2		Run 1	
Process	VBF	Boosted	VBF	Boosted
$Z \to \tau\tau$	1834	11355	34588	20728
$Z \to \ell\ell$	395	1977	1934	868
Top	724	4397	8024	3773
VV	530	5868	2376	1388
Fake	26396	141287	7051	2867
ggF	960	6534	–	39166
VBF	25804	8403	53978	

6.4.3 Impact of Statistics on Training

A large number of training events is a key ingredient for machine learning in particle physics. The number of available events for each physics process is listed in Tab. 6.7. It should be noted, that in Run 2 the number of fake events increased by a large factor, compared to Run 1, where at the same time the number of available signal events decreased. This means the ratio of signal-over-background events dropped to about 0.1 in the Boosted category.

The performance of the BDT is limited by the statistics of the smallest training class, in this case the statistics of the signal processes. To illustrate this, consider the case of the Boosted category, where only 10% of the events are signal events. A classifier, that labels all events as background, has a rate of successful classification of 90%, which could be considered as a good classifier. However, this classifier did not detect a single signal event, and would be of no use in this analysis. In the case of BDTs, the small statistics of the signal class limits the complexity of the boosted decision tree. A boosted decision tree with high complexity learns random fluctuations in the signal class which leads to overtraining and prevents reliable classification. The filtered Monte Carlo production for gluon fusion is used to mitigate the low signal training statistics. The filtered production increases the number of ggF signal samples in the Boosted category by approximately a factor of ten, compare Tables 6.6 and 6.7.

To study the effect of different signal training set sizes, a parameter scan has been performed. The setup of the parameter scan is shown in Tab. 6.8. The usual parameter configurations are duplicated with different signal training set sizes. Each signal event is associated with a random number $r \in [0, 1)$. The inequality

$$r < f \tag{6.11}$$

Table 6.8: Setup of the parameter scan to investigate the impact of statistics of the signal training set.

Parameter	Value(s)
Cross Validation Sets k	10
Signal Training Fraction f	0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1
Input Variables	Run 1 set, see Tab. 6.4
Training Region	regular Boosted
Background (training)	all
Background (validation)	all
Signal (training)	ggF (inclusive + filtered)
Signal (validation)	VBF and ggF (inclusive)
Number of Trees	10, 20, 50, 100, 300, 800, 1000, 1200
Max Depth	10
Shrinkage	0.1, 0.5, 1.0
Min Node Size	0.1, 0.2, 0.5, 1, 2, 5, 10, 20 %

is evaluated for each event, in order to construct a signal training set which holds only the fraction $f \in [0, 1]$ of all signal events. All events that satisfy the condition are included in the set. This means that a set which hold 10% of all events is a subset of the set which hold 20%.

The BDTs are evaluated on the validation set, which uses only the official Monte Carlo production. For each BDT a likelihood fit with the reduced set of systematic uncertainties is carried out, see Sec. 6.2 and Chap. 7. A histogram containing the fit significances is created for each signal training set size. The effect of the training set size is studied by comparing the significance distributions. It is difficult to assess the effect based on the significance distributions. The 10% quantile is derived from the histograms to capture the information about the best performing BDTs for each training set size in a single value. A p-quantile is defined by the fit significance value, which separates the top p BDTs from the rest. Figure 6.14 shows the 10% quantiles as a function of the training set size. The error bars shown in the figure correspond to the 15% and 5% quantiles. It should be noted, that this definition of error bars is arbitrary, and thus the error bars do not correspond to a 1σ confidence level. Different values for p have been tested, however the overall trend of this representation does not depend on the particular value of p.

The data points show that the performance of the top ranked BDTs increases with increasing size of the signal training statistics. The study is however not conclusive, whether this trend extends with even larger training statistics, or if a plateau has already been reached. A deeper investigation would require more Monte Carlo events which are currently not available.

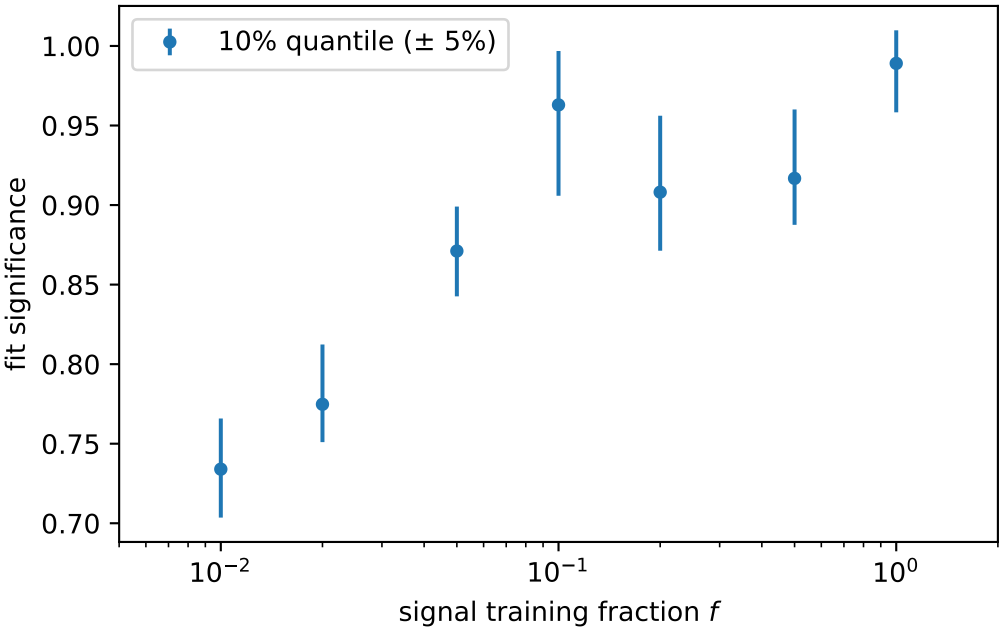

Figure 6.14: Comparison of the BDT performance for different sizes of the signal training set in the Boosted category. The data points show the 10% quantiles. The error bars correspond to the 5% and 15% quantiles and are therefore not 1σ-error bars.

6.5 Analysis of k-fold Cross Validation

In Sec. 6.1.6 the principle of k-fold cross validation was introduced. This principle with $k = 10$ is applied for all studies in this thesis unless otherwise noted. However, the choice of $k = 10$ is arbitrary. The study presented in this section investigates the effect of different values of k. With increasing k, the training set size converges towards the total amount of Monte Carlo events. Based on Sec. 6.4, one can expect that the performance increases also with increasing k. The extreme case of $k = n$ is used in some fields of machine learning, where n denotes the number of training observations. This would mean, that there is one boosted decision tree for each event. The application of the technique $k = n$ for this thesis is computationally not feasible in a reasonable amount of time. It is therefore advisable to limit the number k of cross validation sets. Since the fraction t of training events approaches the total number of events with

$$t = 1 - \frac{1}{2k} \tag{6.12}$$

it can be assumed that the performance also approaches a limit and that the gain at high k values is marginal.

A parameter scan has been performed for the present study. A new axis has been introduced alongside the regular variable axes in the parameter grid. Different values for k are tested to study the dependence on the number of cross validation sets. Similarly to the evaluation procedure in the previous study, the BDTs from the parameter scan are used on the validation set. A likelihood fit with a reduced set of

Table 6.9: Setup of the parameter scan to investigate the importance of different numbers of k in cross validation sets.

Parameter	Value(s)
Cross Validation Sets k	3, 5, 7, 10, 13, 16, 20
Input Variables	Run 1 set, see Tab. 6.4
Training Region	regular VBF
Background (training)	all
Background (validation)	all
Signal (training)	VBF only
Signal (validation)	VBF and ggF (incl.)
Number of Trees	10, 20, 50, 100, 200, 400, 800, 1000
Max Depth	10
Shrinkage	0.03, 0.1, 0.3, 0.8
Min Node Size	0.1, 0.2, 0.5, 1, 2, 5, 10, 20 %

systematic uncertainties is used to derive the expected significances, see Sec. 6.2 and Chap. 7. The significance values are filled in separate histograms, one histogram for each value of k. As described in Sec. 6.4.3 the 10% quantile is calculated to represent the performance of the boosted decision trees in a single value. The results are shown in Fig. 6.15.

From Fig. 6.15 one might read a slight improvement in BDT performance for increasing k values, however, the question remains, whether this improvement is significant and not only attributable to statistical fluctuations. This study does not show that choosing lower or higher k values is beneficial for the analysis. In cases where computational resources are limited, the choice $k = 3$ seems valid. However, for the remainder of this thesis, the choice of $k = 10$ is unaltered.

6.6 Boosted Decision Tree Optimization

A central component of this thesis is a comparison between the multivariate analysis and the cut-based analysis. The cut-based analysis has been optimized and studied intensively in 2016 by the ATLAS HLeptons analysis group. The goal of this section is to study and optimize the training parameters of the multivariate analysis. The first two sections study the performance depending on different training regions and whether it is beneficial to include the fake background in the training. The last section is about the selection of the boosted decision trees used for the multivariate analysis. The studies presented in this section all share the same large scale parameter scan. The setup of this parameter scan is outlined in Tables 6.10 and 6.11. The complexity of a tree in the boosted decision tree is controlled by the parameter minimum node size. Various different values for this parameter are tested in the parameter scan. The depth of a tree in the boosted decision tree is set to a high value, such that the depth is only limited by the condition on the minimal node size. This offers more flexibility. Individual paths in the boosted decision tree can extend to

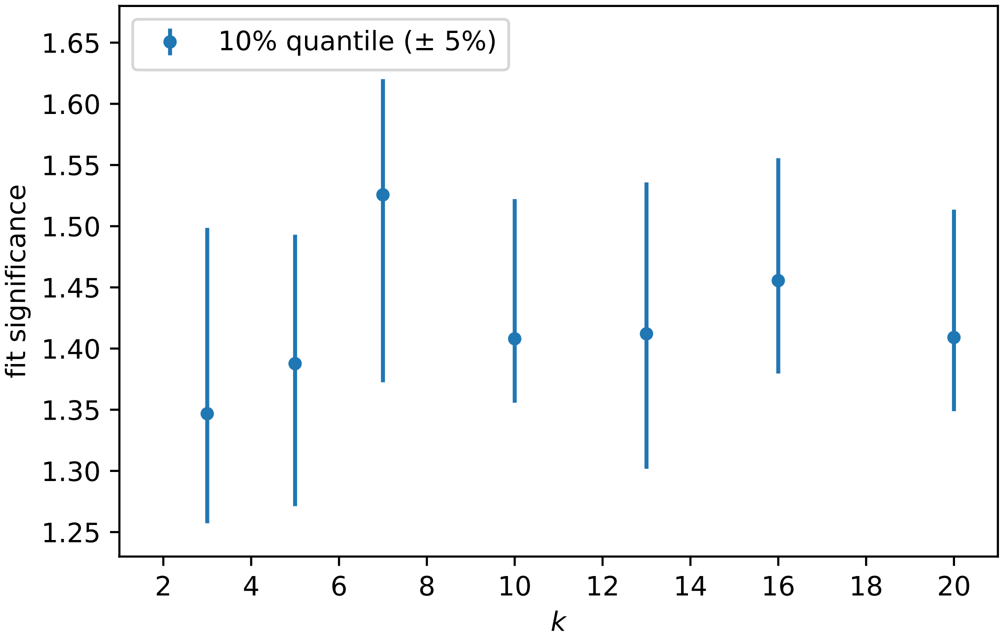

Figure 6.15: Comparison of the BDT performance for different numbers k of cross validation sets in the VBF category. The data points show the 10% quantiles. The error bars correspond to the 5% and 15% quantiles and are therefore not 1σ-error bars.

larger depth if there is still enough statistics. The parameter scan also features the two axes *number of trees* and *shrinkage*. Both interact with the boosting behavior and control the complexity of the whole boosted decision tree.

6.6.1 Training Region

In Run 1 the training of the BDTs had been performed in regions looser than the regular MVA regions. The enlarged regions have only been used for the training. Similarly to the Run 1 procedure, the likelihood fit to evaluate the BDT on the validation set is performed in the regular MVA regions. The motivation behind this is, that by enlarging the training regions, the training can benefit from an increase in statistics. Conversely, if a BDT is trained in an enlarged phase space, it might excel in separating signal from background in regions of phase space which are not relevant for the analysis and perform rather badly in the regular MVA regions.

The effect on the training region is studied using the large scale parameter scan introduced in the previous section. The parameter scan features a variable axis, which specifies the training region. Following the suggestions from Run 1, the training for VBF is done in the regular region and repeated in regions with the cuts $m_\mathrm{T} < 70\,\mathrm{GeV}$ and/or $\Delta\eta^{jj} > 3$ dropped. Similarly, the training for Boosted is performed in the regular regions and in an enlarged region without the $m_\mathrm{T} < 70\,\mathrm{GeV}$ cut. The BDTs are evaluated on their validation sets. The expected significance is derived using a likelihood fit with a reduced set of systematic uncertainties, see Sec. 6.2 and Chap. 7. The fit significances are filled into separate histograms de-

Table 6.10: Setup of the parameter scan to optimize the multivariate analysis in the VBF category. The same parameter scan is used to study the impact of the training regions, whether it is beneficial to include fake events in the training and to select the boosted decision trees for this analysis.

Parameter	Value(s)
Cross Validation Sets k	10
Input Variables	Run 1 set, see Tab. 6.4
Training Region	VBF: regular, w/o m_{T}, w/o $\Delta\eta$, w/o $\Delta\eta$ and m_{T}
Background (training)	all, w/o fakes
Background (validation)	all
Signal (training)	VBF only
Signal (validation)	VBF and ggF (incl.)
Number of Trees	10, 20, 50, 100, 200, 400, 800, 1000, 1200
Max Depth	15
Shrinkage	0.03, 0.1, 0.3, 0.8, 1
Min Node Size	0.1, 0.2, 0.5, 1, 2, 5, 10, 20 %

Table 6.11: Setup of the parameter scan to optimize the multivariate analysis in the Boosted category. The same parameter scan is used to study the impact of the training regions, whether it is beneficial to include fake events in the training and to select the boosted decision trees for this analysis.

Parameter	Value(s)
Cross Validation Sets k	10
Input Variables	Run 1 set, see Tab. 6.4
Training Region	Boosted: regular, w/o m_{T}
Background (training)	all, w/o fakes
Background (validation)	all
Signal (training)	ggF (incl. + filt.)
Signal (validation)	VBF and ggF (incl.)
Number of Trees	10, 20, 50, 100, 200, 400, 800, 1000, 1200
Max Depth	15
Shrinkage	0.03, 0.1, 0.3, 0.8, 1
Min Node Size	0.1, 0.2, 0.5, 1, 2, 5, 10, 20 %

pending on the training region. The resulting histograms are shown in Fig. 6.16.

The distributions for VBF and Boosted are derived from a grid scan, where the $\Delta\eta^{jj}$ cut is dropped show a peak at lower significance values. Also, looking at the high tails of these two distributions, it is clear that a training region without the angular cut, is not beneficial for the sensitivity of the analysis. For VBF and Boosted, there seem to be only minimal differences between the two distributions derived from parameter scans in the regular regions and in regions enlarged by dropping the m_T cut. This statement is true for the position of the mean and the extent of the high tails.

The conclusion of this study is that, in terms of sensitivity the cut $\Delta\eta^{jj} > 3$ should be applied during training. On the other hand, no general statement about the $m_\mathrm{T} < 70\,\mathrm{GeV}$ cut is possible. Whether it is beneficial to apply this cut during training, depends on the other training parameters.

6.6.2 Fake Treatment

A similar study has been carried out to investigate whether it is beneficial to include fake backgrounds in the training. The study follows the prescriptions given in Sec. 6.6.1. The parameter scan is performed for different background compositions. One variant is to use all background processes for the training, the other is to exclude fake background during training.

The resulting histograms are shown in Fig. 6.17. In both categories the bulk of the BDTs seem to scatter at the same values of expected significances. In the VBF category the distributions are asymmetric. BDTs trained with fake background have a more pronounced tail at high sensitivity values. In Boosted, however the difference between the two distributions is smaller and no general statement is possible.

6.6.3 BDT Selection

There are several different methods to select a BDT after a parameter scan has been performed. One method is to search for a region in parameter space, where good BDTs accumulate. Since the parameter space is a multi-dimensional space, its graphical representation is not easy. One method could be to choose two variables as the axes of a plot, and project the remaining space onto these two axes by averaging over all other axes in the parameter grid. The average fit significance can then be represented with a color map. It is possible that the topology in parameter space could be rather complicated. In that case averaging over many dimensions in parameter space, could hide important structures or even lead to false interpretations. This method has been tested at an early stage of the analysis, but no conclusive trend, or sweet spot in parameter space could be identified.

In this thesis a simplified approach is pursued, to avoid the complications mentioned above. The final BDTs, which are used to produce the distributions used in the final likelihood fit of the analysis, are selected using a ranking of all BDTs by their significance estimate on the validation set. This approach simply selects the best BDT for VBF and for Boosted. As discussed in Sec. 6.1.5, it might happen that the selected BDTs perform exceptionally good on the validation set by chance. In that case, the selected BDTs could lead to suboptimal sensitivities of the analysis. The selection is based on the performance on the validation set, therefore this

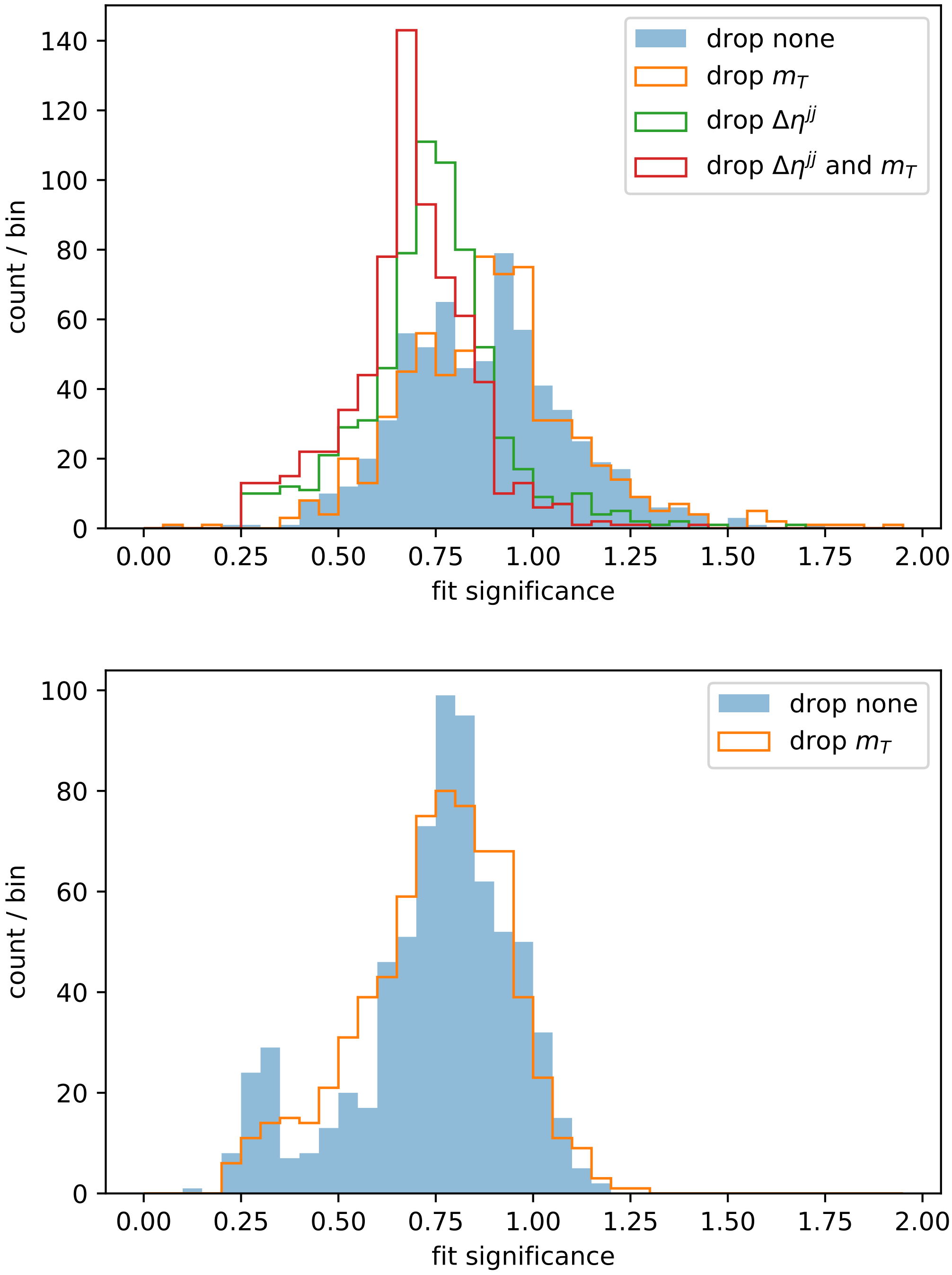

Figure 6.16: Comparison of BDT performance on the validation set for different training regions defined by dropping selection criteria in the VBF category (top) and Boosted category (bottom).

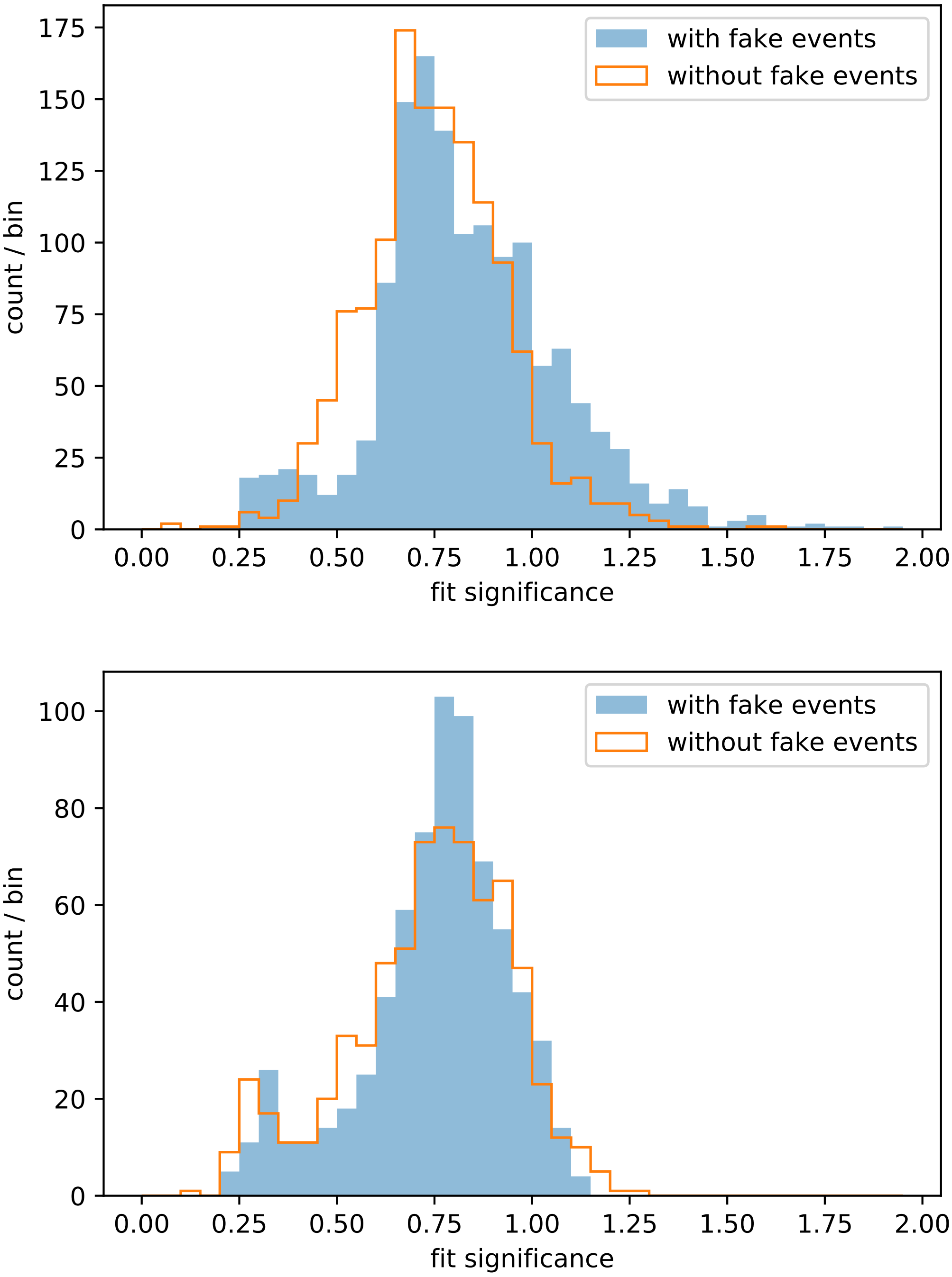

Figure 6.17: Comparison of BDT performance on the validation set for different background compositions including or excluding events with a fake tau in the VBF category (top) and the Boosted category (bottom).

Table 6.12: Parameter listing of the best BDT training configurations.

	VBF	Boosted
number of trees	50	200
minimum node size	1	1
maximal depth	15	15
shrinkage	0.1	0.3
region	drop m_{T} cut	regular
background	with fakes	without fakes
fit significance	1.94	1.19

procedure can not lead to a biased BDT on the test set used in the analysis, which means that the final result of the analysis is still valid even if the selected BDT are outliers.

The BDT training parameter configurations in each category with the highest[3] fit significance on the validation set are listed in Tab. 6.12. If these numbers are compared to the training parameters in Run 1 as listed in Ref. [15], one can see, that in Run 1, the number of trees was higher by a factor of eight and three for VBF and Boosted, respectively. The shrinkage parameters are similar to the configuration in Run 1. The depth and minimal node size parameters can not be directly compared, because in Run 1 the complexity of the tree was limited by a maximum tree depth as an optimization meta-parameter, while the minimum node size was set to very small fixed values. In this thesis, the depth is limited by the minimal node size and the depth is set to a high value to give the BDT training a greater flexibility. The variable ranking of these two BDTs is shown in Tables 6.13 and 6.14.

6.7 BDT Output and BDT Validation

After a BDT training configuration has been selected for both categories, the BDTs can be applied to their test sets. The BDT score distributions produced on the test sets is used as the discriminating variable in the multivariate analysis. The distributions for both categories are shown in Fig. 6.18. The data entries are blinded in both distributions for high BDT scores. All data points are removed from the histogram above a BDT score of 0.4. The VBF BDT shows a strong separation between background and signal. The fact that the background and signal do not peak at -1 and 1 is due the relatively low number of trees and shrinkage value. This is, however, not a problem, since the BDT score itself is an arbitrary measure. The output distribution in the Boosted category shows also good separation between background and signal. Both distributions show good agreement between data and Monte Carlo in the unblinded range.

[3] Technical difficulties were experienced when the best BDTs were used in the analysis stemming from large memory consumption, due to a large number of very deep trees in the BDT. If such difficulties occurred, the BDT is ignored, and the next best BDT configuration is used. This means the term *best* actually refers to the *best without technical issues.*

Table 6.13: Variable ranking for the selected boosted decision tree in the VBF category as shown in Tab. 6.12.

Variable	Importance
$\ell\,\eta$ centrality	0.2027
$E_{\mathrm{T}}^{\mathrm{miss}}\,\phi$ centrality	0.1983
m_{MMC}	0.1499
m^{jj}	0.1311
$\Delta\eta^{jj}$	0.1000
m_{T}	0.0679
$p_{\mathrm{T}}^{\mathrm{total}}$	0.0677
$\Delta R^{\mathrm{lep\,had}}$	0.0625
$\eta^{j_0}\cdot\eta^{j_1}$	0.0195

Table 6.14: Variable ranking for the selected boosted decision tree in the Boosted category as shown in Tab. 6.12.

Variable	Importance
$\Delta R^{\mathrm{lep\,had}}$	0.2944
m_{MMC}	0.2018
m_{T}	0.1295
$p_{\mathrm{T}}^{\mathrm{lep}}/p_{\mathrm{T}}^{\mathrm{had}}$	0.1322
scalar $\sum p_{\mathrm{T}}$	0.1169
$E_{\mathrm{T}}^{\mathrm{miss}}\,\phi$ centrality	0.1249

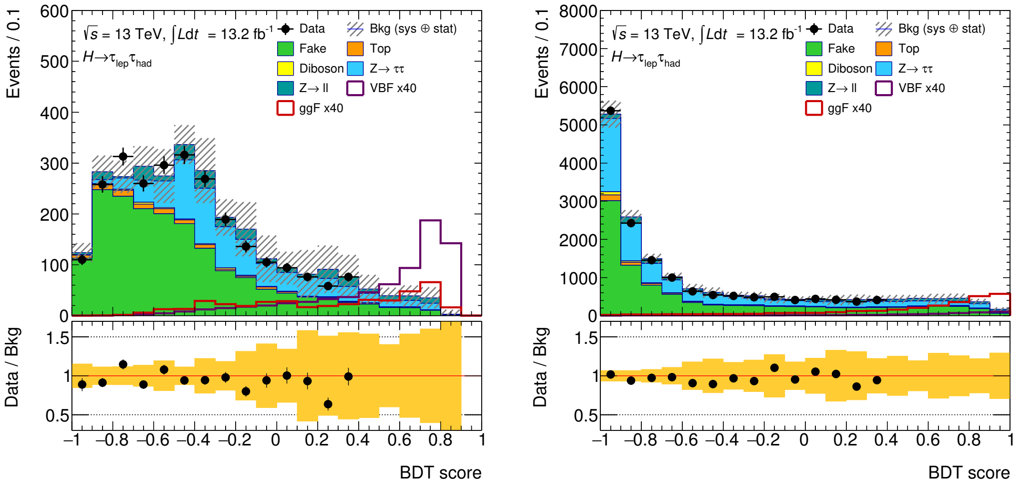

Figure 6.18: Boosted decision tree output score distributions in VBF (left) and Boosted (right) signal regions.

In order to validate the boosted decision trees, they have been applied to the control regions as defined in Sec. 5.7. Figure 6.19 shows the BDT output score distributions in the VBF categories. Figure 6.20 shows them in the Boosted categories. Since the signal contribution is negligible in the control regions, no blinding criteria has been applied to these plots. The distributions show good agreement between data and Monte Carlo in the control regions given the statistical and systematic uncertainties.

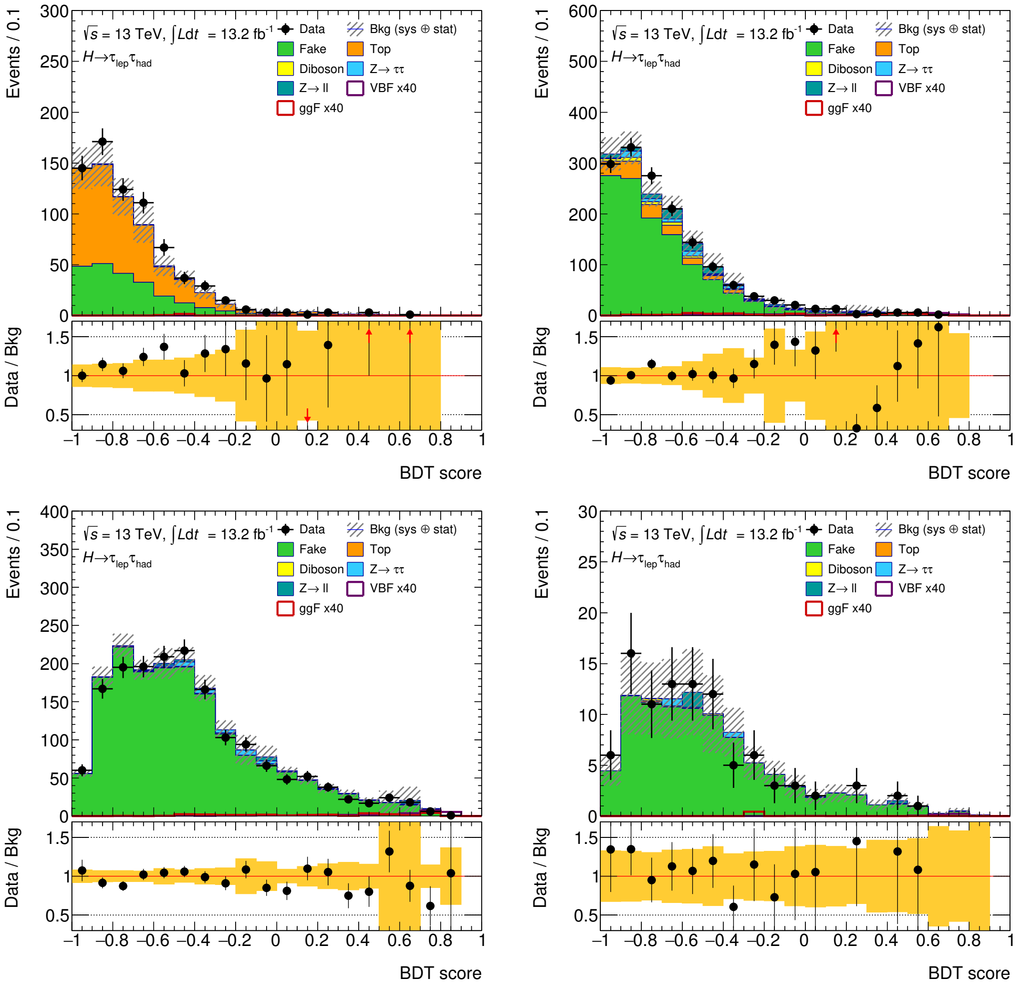

Figure 6.19: Boosted decision tree output score distributions in the VBF control regions. The top row shows the distributions in the top (left) and W (right) control regions. The bottom row shows the distributions in the QCD (left) and Z (right) control regions.

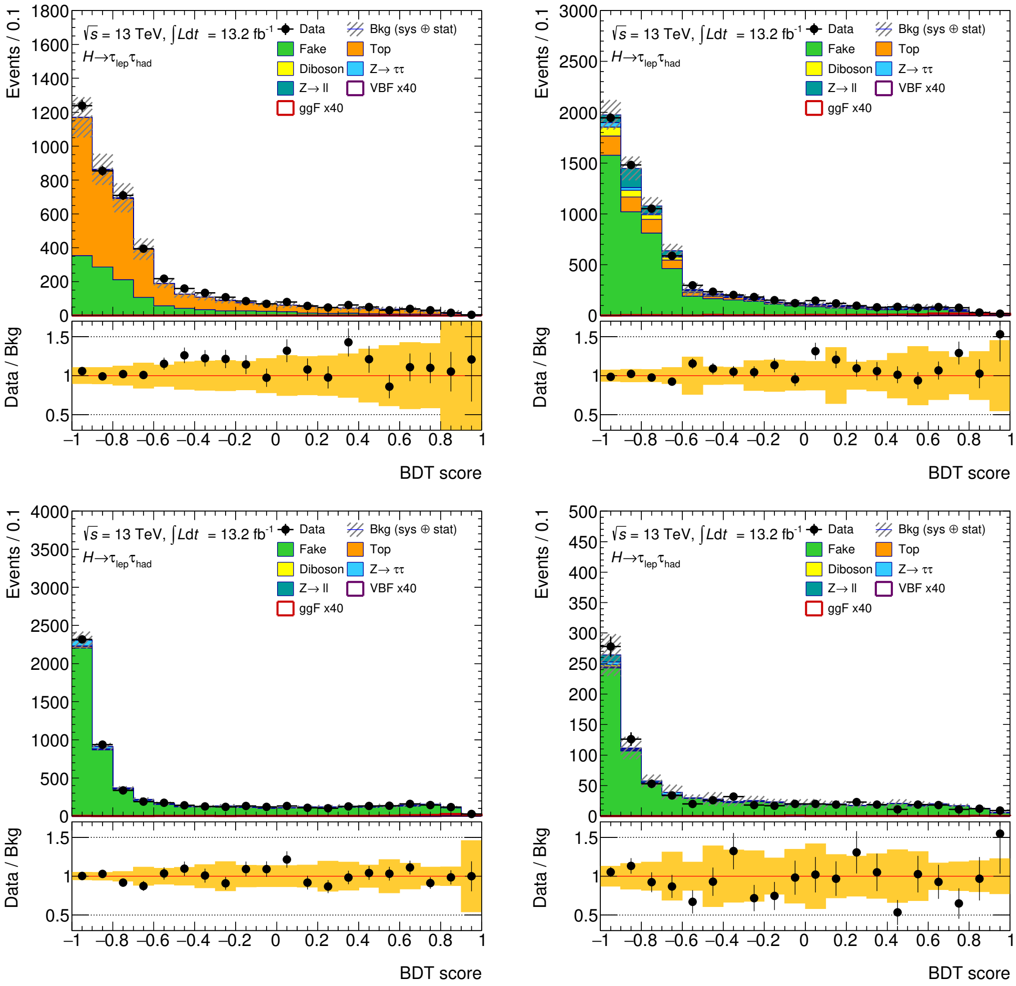

Figure 6.20: Boosted decision tree output score distributions in the Boosted control regions. The top row shows the distributions in the top (left) and W (right) control regions. The bottom row shows the distributions in the QCD (left) and Z (right) control regions.

CHAPTER 7

Signal Strength Extraction

The signal strength μ is defined as the ratio of the observed $H \to \tau\tau$ yield over the Standard Model prediction. The signal strength for both analysis methods is extracted with a likelihood fit. The overall fit is carried out for the cut-based analysis and the multivariate analysis independently. Since this analysis is not approved by the ATLAS collaboration, the final fit is not using real data events. Instead pseudo data yields and distributions are built from the background and signal model by using the expected values assuming the validity of the SM. This is commonly referred to as an *Asimov* fit. The first section in this chapter starts by introducing all systematic variations used for the cut-based analysis and the multivariate analysis. Secondly, this chapter outlines the fit model used in this thesis. The chapter closes with a section discussing the results from both fits and thus compares the outcomes of the two analysis strategies.

7.1 Systematic Uncertainties

Systematic uncertainties play a crucial role when it comes to the sensitivity of the analysis. They enter the analysis in many different ways. Most of the systematic uncertainties quantify a lack of precise knowledge about the detector response, the detector calibration or enter as reconstruction and identification artifacts [19]. The systematic uncertainties are considered in the analysis by adhering to the following procedure.

Each systematic uncertainty describes the uncertainty of a specific parameter of the simulation, reconstruction, calibration, event generation, parton showering, parton distribution function or another components of the analysis. Each event is reevaluated for each systematic uncertainty under varied conditions. This means the events are reprocessed by taking the $+1\sigma$ and -1σ variations of the parameter under consideration. The distributions that enter the fit are available in multiple version, one version for each systematic variation. This can be viewed as a numerical approximation of partial derivatives. The distributions for the up and down variations quantify the change of the distribution by varying the parameter up and down by one standard deviation. Since simultaneous variations of two parameters are not considered, this procedure is only correct if the variations are not physically correlated.

Systematic variations can usually be classified in two categories: weight systematic uncertainties and kinematic systematic uncertainties. The former kind affects only the event weight. These systematic uncertainties do not change the kinematic variables of an event. It is not necessary to reprocess the event, since the nominal event can be reused with a different weight. The latter kind of systematic variations are computationally more difficult, because this kind of systematic uncertainty changes the kinematic variables of the event. All the events have to be reprocessed for each kinematic variation.

Looking from a different perspective, most systematic uncertainties can also be classified into three categories: systematics affecting the resolution of the measurement, the scale of the momentum or energy measurement, or the efficiency to select or reject individual objects (leptons, taus, jets) and therefore to select or reject the event.

The application of the systematic uncertainties follows the ICHEP 2016 recommendations of each of the ATLAS combined performance groups. In the following the experimental systematic uncertainties are listed separately for these three categories. The list is split into several sections depending on the object the systematic uncertainty concerns. Some systematic uncertainties stem from a statistical fluctuation, but since they affect the analysis as a whole, they are turned into systematic uncertainties.

7.1.1 Muon Uncertainties

Systematic uncertainties related to muons enter the analysis via the muon from the leptonic τ decay. Trigger efficiencies change the total yield of the muons. Scale and resolution uncertainties have an impact on m_{MMC}, since it is calculated using the momentum of the lepton. The following list describes all systematic uncertainties stemming from muons.

Resolution uncertainties of the inner detector tracks (`MUONS_ID`, up/down) and muon spectrometer tracks (`MUONS_MS`, up/down).

Scale uncertainties of the muon momentum measurement (`MUONS_SCALE`, up/down)

Efficiency uncertainties originate from different components. The Muon triggers used in 2015 and 2016 consist of systematic and statistical uncertainties (`MUON_EFF_Trig[Syst|Stat][2015|2016]`, up/down). The uncertainties on the identification efficiencies are composed of a statistical and a systematic part (`MUON_EFF_[STAT|SYS]`, up/down). Additionally there are the uncertainties stemming from the isolation requirement (`MUON_ISO_[STAT|SYS]`, up/down) and from the track-to-vertex association (`MUON_EFF_TTVA_[STAT|SYS]`, up/down).

7.1.2 Electron Uncertainties

The systematic variations for electrons follow the single-nuisance-parameter scheme ("total"). The systematic uncertainties of the electrons enter the analysis in a similar way as in the case for muons. Additionally, electrons, unlike muons, can be falsely reconstructed as τ_{had}. The m_{MMC} calculation with wrong particle settings, results in incorrect results. The background from misidentified electrons peaks in the m_{MMC} range close to the signal region, see Figures 5.2 and 5.3. This means that a variation of the electron related systematic uncertainties are likely to have a large effect on the signal strength. The following list describes systematic uncertainties stemming from electrons in the analysis.

Resolution uncertainty of electrons (EG_RESOLUTIONS_ALL, up/down).

Scale uncertainties of the energy of electrons have several independent sources stemming from different detector components (EG_SCALE_E4SCINTILLATOR, EG_SCALE_LARCALIB_EXTRA2015PRE and EG_SCALE_LARTEMPERATURE_EXTRA[2015|2016]PRE, all up/down). Other uncertainties affecting the energy scale of the electrons are combined into a single uncertainty (EG_SCALE_ALLCORR, up/down).

Efficiency uncertainties are introduced for different components of the analysis. Systematic variations are calculated for the uncertainties on the trigger efficiency (EL_EFF_TRIG_TOTAL), on the reconstruction efficiency (EL_EFF_RECO_TOTAL), on the identification efficiency (EL_EFF_ID_TOTAL) and on the efficiency of the isolation requirement (EL_EFF_ISO_TOTAL).

7.1.3 Tau Uncertainties

The systematic uncertainties related to hadronic τ leptons affect the analysis in a similar way as in the case of muons. The following list describes systematic uncertainties stemming from taus in the analysis.

Scale uncertainties affecting the energy scale, stemming from the detector (TAU_TES_DETECTOR, up/down), in-situ measurements (TAU_TES_INSITU, up/down) and modeling and closure tests (TAU_TES_MODEL, up/down).

Efficiency uncertainties are introduced from different components of the analysis. Systematic variations are calculated for the uncertainties on the trigger efficiency (TAU_TRIG_STAT[DATA|MC] and TAU_TRIG_[SYST|TOTAL2016], both up/down), the tau reconstruction efficiency (TAU_EFF_RECO_[TOTAL|HIGHPT], up/down), the electron overlap removal (TAU_EFF_ELEOLR_TRUE[ELECTRON|HADTAU], up/down) and the jet identification/rejection (TAU_EFF_ID_[TOTAL|HIGHPT], up/down).

7.1.4 Jet Uncertainties

Systematic uncertainties of jets affect the analysis in two ways. The selection cuts for the VBF category depend directly on jet related quantities. Since the Boosted region depends on an event failing the VBF criteria, the Boosted category is indirectly affected by these systematic uncertainties. Variations in the energy scale or resolution can lead to migrations between the two categories. A second mechanism how jet related uncertainties enter the analysis is via $E_{\mathrm{T}}^{\mathrm{miss}}$. The missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$ is the momentum needed to balance all visible parts of the event. Changes in jet related quantities affect the momentum necessary to balance these jets. The analysis is affected by these systematics, because $E_{\mathrm{T}}^{\mathrm{miss}}$ is included in the m_{MMC} calculation. The following list describes systematic uncertainties stemming from jets in the analysis.

Resolution uncertainties for the jet energy are combined in a single nuisance parameter (`JER`, one-sided)

Scale uncertainties affecting the jet energy scale have several different sources. The general systematics affecting the jet energy scale are combined in 6 effective nuisance parameters (`JES_EffectiveNP_[1-6]`). More specific systematic variations are introduced to consider effects related to b-jets (`JES_BJES_Response`, up/down), an η-dependence (`JES_EtaInter_Model`, `JES_EtaInter_NonClosure` and `JES_EtaInter_TotalStat`, all up/down), the particle flavors (`JES_Flavor_Comp` and `JES_Flavor_Resp`, both up/down), pile-up (`JES_PU_OffsetMu`, `JES_PU_OffsetNPV`, `JES_PU_PtTerm` and `JES_PU_RhoTopology`, all up/down) and other effects (`JES_PunchThrough_MC15` and `JES_SingleParticle_HighPt`, both up/down).

Efficiency uncertainties cover two main uncertainties. Firstly a systematic uncertainty for the jet vertex tagger (`JVT`, up/down) is introduced. Secondly efficiency uncertainties for flavor tagging are introduced. In total 14 effective systematic variations are introduced related to b tagging (`btag_b_[0-2]`, `btag_c_[0-3]`, `btag_light_[0-4]` and `btag_extrapolation[_from_charm]`, all up/down).

7.1.5 Missing Transverse Energy Uncertainties

Since $E_{\mathrm{T}}^{\mathrm{miss}}$ describes missing transverse energy, $E_{\mathrm{T}}^{\mathrm{miss}}$ is also affected by other systematic uncertainties, such as the uncertainties stemming from jets. The systematics specifically for $E_{\mathrm{T}}^{\mathrm{miss}}$ stem from jets below a certain threshold, which are not considered as proper jets. These *soft tracks*, however, contribute to the missing transverse energy. The following list describes systematic uncertainties stemming from $E_{\mathrm{T}}^{\mathrm{miss}}$ in the analysis.

Resolution uncertainties related to missing transverse energy (`MET_SoftTrk_ResoPara` and `MET_SoftTrk_ResoPerp`, both one-sided)

Scale uncertainties affecting the missing transverse energy (`MET_SoftTrk_Scale`, up/down)

7.1.6 Pile-Up Uncertainty

As explained in Sec. 4.4 pile-up events are simulated by mixing the detector response with simulated pile-up events. The Monte Carlo has to be reweighted to match the observed pile-up profile. A nuisance parameter (PRW_DATASF, up/down) is introduced to cover the uncertainties of this procedure. The kinematic variables should not depend on the pile-up conditions. Such a dependence might manifest itself as an impact of this systematic uncertainties.

7.1.7 Luminosity Uncertainty

The uncertainties of the luminosity for the data set used in this analysis has been determined by ATLAS luminosity working group using the methodology described in Ref. [32]. The uncertainty on the luminosity measurement is 2.9 % and can be considered uncorrelated between 2015 and 2016. The luminosity uncertainties act as an overall weight uncertainty.

7.1.8 Background Model Uncertainties

The background model for events including jets faking taus is detailed in Sec. 5.3.5. The statistical uncertainties $\sigma_N = \sqrt{N}$ of the yields N in Eq. (5.11) affect the analysis as systematic uncertainties. The statistical uncertainties are therefore propagated to the fake factor f as σ_f with Gaussian error propagation. Systematic variations are then calculated for $f \pm \sigma_f$. To account also for uncertainties of R_{QCD}, a variation of R_{QCD} by $\pm 50\%$ is introduced. However, the variation of 50% is arbitrary. The other R_i are scaled to satisfy the equation $\sum_i R_i = 1$. Both variations, the statistical variation of f and the variation of R_{QCD} are combined to form a single nuisance parameter (lh_fake, up/down).

Theory uncertainties of the Z+jets backgrounds are derived using systematic variations from Monte Carlo samples generated with Sherpa. To apply the variation to Madgraph samples, the relative variation of the corresponding Sherpa sample is determined. These variations are associated with the nuisance parameters for the factorization scale (Theo_Ztt_fac, up/down) and the renormalization scale [19] (Theo_Ztt_ren up/down).

7.1.9 Signal Modeling Uncertainties

The uncertainty on the Higgs production cross-section is taken from [21] assuming the Higgs mass $m_H = 125.09\,\text{GeV}$. The QCD scale uncertainties are assumed to be 4.0% for gluon fusion and 2.1% for vector boson fusion. The cross-section uncertainties stemming from the Parton Distribution Function (PDF) uncertainties are $\pm 3.3\%$ and $\pm 2.2\%$ for ggF and VBF respectively. The uncertainty of the branching ratio of $H \to \tau\tau$ is $^{+1.17\%}_{-1.16\%}$.

7.2 Fit Model

The signal strength μ is extracted with a binned likelihood fit taking all systematic uncertainties into account. The fit model is similar for the cut-based analysis and the multivariate analysis in order to achieve comparability. In the following a simplified description of the fit model is given. The details of the fit model and the fitting algorithms are beyond the scope of this thesis.

The input of the fit is a collection of histograms. The histograms in the signal regions show the distribution of the discriminating variable. In the case of the cut-based analysis this is m_{MMC}, in case of the multivariate analysis this is the boosted decision tree output score. In the cut-based analysis the signal regions are VBF tight and loose, and Boosted high and low. In the multivariate analysis the signal regions are the inclusive VBF and Boosted regions. The top control region is used to constrain the normalization of top processes, therefore a histogram with only a single bin corresponding to the total yield is used in the control region.

The fit uses Asmiov data, which means that the nominal background plus signal expectation is used as data. The fit input consists of one histogram for each signal region and control region with these pseudo data distributions. The input includes also histograms for the background and signal estimation. This means that, similarly to the data histograms, one histogram for each region is used with the nominal background distributions. The same set of histograms is included in the input for the expected signal events.

In order to take systematic uncertainties into account, the set of histograms of the background and signal is duplicated for each systematic variation (up and down). For each systematic uncertainty i a free, continuous parameter θ_i is introduced, commonly referred to as *nuisance parameter*. A technique called *moment morphing* is employed to interpolate between the $\pm 1\sigma$ variations. A non-zero nuisance parameter means that the distributions of the background and signal models are shifted towards the corresponding systematic variation. All nuisance parameters are combined into the vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots)$. The background and signal models are therefore available as a function of the nuisance parameters $\boldsymbol{\theta}$.

Finally the signal strength parameter μ is defined. The parameter μ scales the contribution of the signal. The value of $\mu = 1$ corresponds to the Standard Model expectation. The choice $\mu = 0$ corresponds to the background-only hypothesis.

The heart of the fit is a likelihood function L which simultaneously quantifies the likeliness of two aspects. Firstly it quantifies how likely it is to have a certain choice of $\boldsymbol{\theta}$ given the initial systematic uncertainties. Secondly L quantifies how likely it is to get the data observation from a random experiment given the statistical uncertainties of the background and signal model. For technical reasons it is beneficial to consider the negative logarithm of the likelihood function (NLL). The fit procedure consists of minimizing the NLL function (maximizing the likelihood) by varying the parameters $\boldsymbol{\theta}$ and μ.

By minimizing the NLL function the fit finds the most probable values of the parameters $\boldsymbol{\theta}$ and μ . The curvature (*i.e.* the second derivatives) of the NLL with respect to the parameters $\boldsymbol{\theta}$ indicate how strong this analysis is able to constrain the systematic variations. Since each systematic variation has an initial uncertainty, one is able to compare the initial uncertainty with the uncertainty derived from the fit. It can be a sign of a problem in the analysis, if the fit constrains the nuisance parameters more than what was expected from the pre-fit systematic uncertainties.

Another important test is to check how strong the best μ value depends on the variation of a systematic nuisance parameters θ_i. The systematic variations can be ranked by their impact on μ. This gives valuable information how to improve the sensitivity of the analysis.

The minimization of NLL is performed twice. The first round is an unconditional

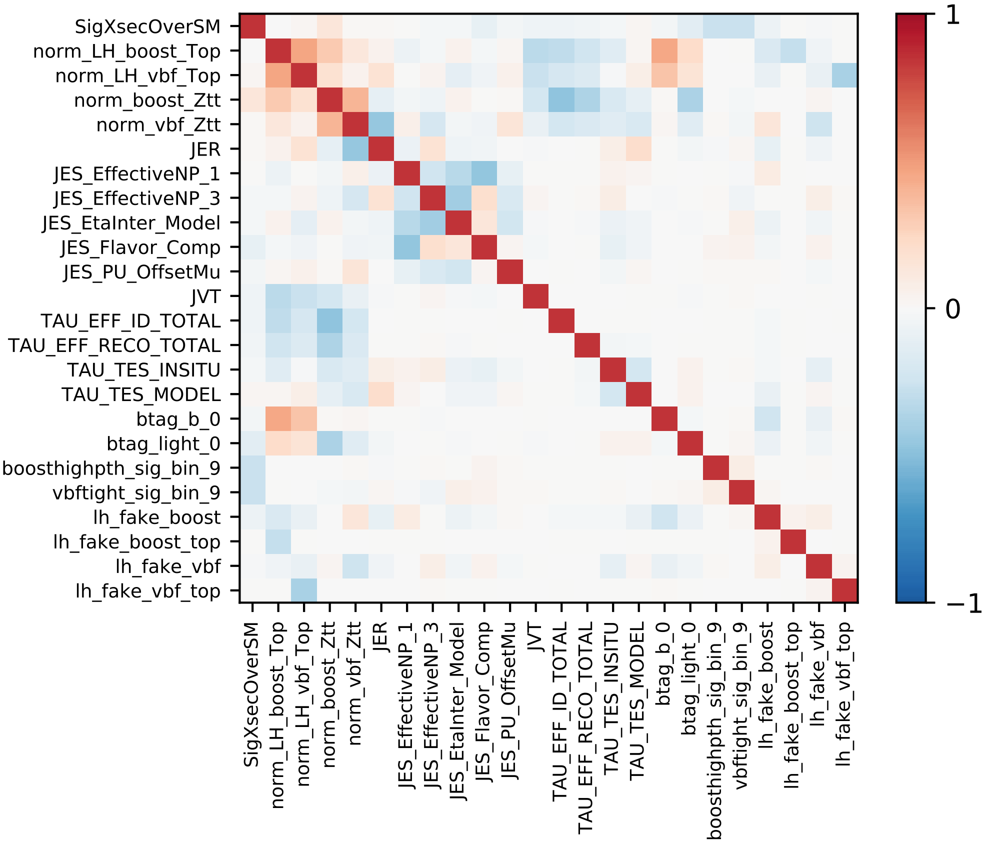

Figure 7.1: Linear correlation coefficients between the systematic variations for the cut-based analysis. Systematic uncertainties with absolute correlations coefficients below 25% are pruned away for this visualisation. The signal strength μ is denoted by `SigXsecOverSM`.

fit, what means that parameters $\boldsymbol{\theta}$ and μ are free and can both be varied during the minimization. For the second round, the signal strength is constrained to $\mu = 0$, which corresponds to the background only hypothesis. By comparing the two minima of the NLL, one can derive the significance, which measures the probability that the observed distributions is not random fluctuation of the background-only hypothesis. Let L_{min} denote the minimal value of the likelihood function L from the unconditional fit and L_{cond} denote the minimal value from the conditional fit with $\mu = 0$. The significance Z is then defined as

$$Z = \sqrt{2 \cdot (\log L_{\text{min}} - \log L_{\text{cond}})}. \tag{7.1}$$

7.3 Fit Result

The fit described in the previous section was performed separately for the cut-based analysis and the multivariate analysis.

As described in Sec. 7.1, it is assumed that the systematic variations are not statistically correlated. Figures 7.1 and 7.2 shows the linear correlation coefficients between the effect of the systematics variations as a color map. In this context the distinction between causation and correlation is important. The effects of two

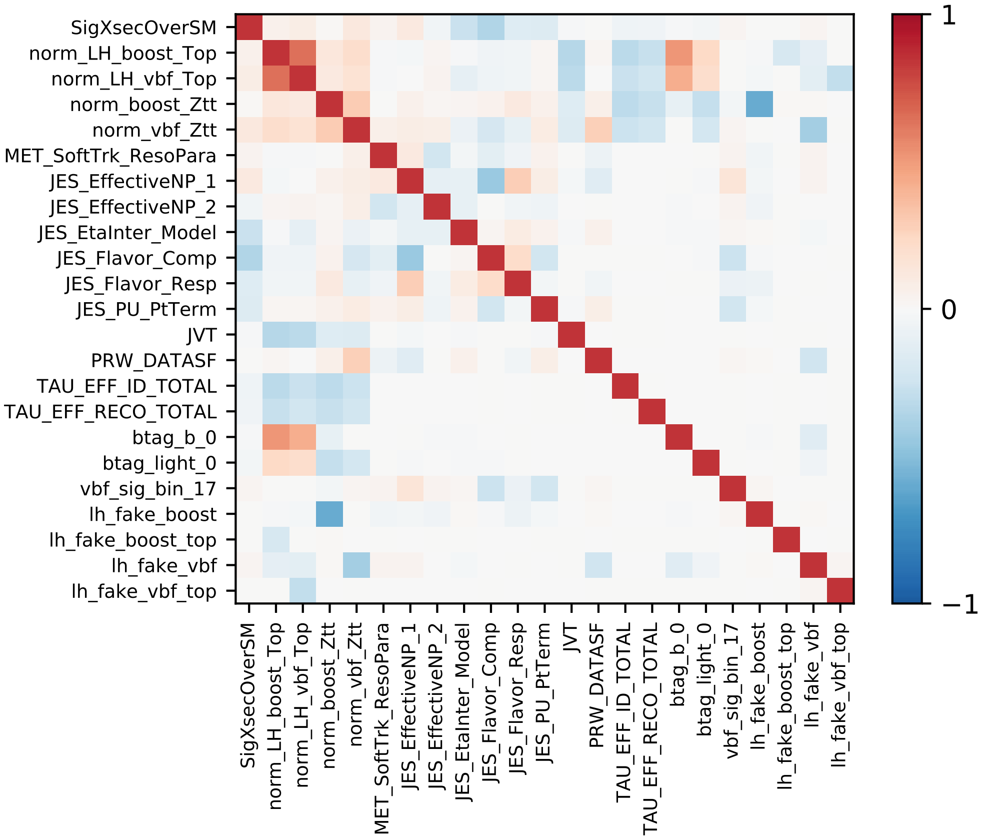

Figure 7.2: Linear correlation coefficients between the systematic variations for the multivariate analysis. Systematic uncertainties with absolute correlations coefficients below 25% are pruned away for this visualisation. The signal strength μ is denoted by `SigXsecOverSM`.

systematic variations can be linearly correlated, even though they are statistically independent[1]. The correlation plots in Figures 7.1 and 7.2 give insight into the analysis and offers a cross check of the systematic variations.

For both CBA and MVA the top and Z normalizations are positively correlated between Boosted and VBF. It seems questionable how this correlation can occur across the categories, since each has an independent normalization. However, if the normalization is changed in one category, other nuisance parameters adjust, to compensate for this, which then also affects the other category.

The normalization factors for top and Z are stronger correlated to the systematic uncertainties of τ identification and reconstruction efficiencies for the cut-based analysis than for the multivariate analysis. It seems quite natural that a variation of the τ identification and reconstruction efficiency affects the normalization of background processes with a τ. Furthermore, both analyses show a correlation between the top normalization and systematic variations related to b-tagging. This is expected, since the b-veto cut is applied to remove background from $t\bar{t}$ events. In addition, in both analysis methods the systematic uncertainties related to the jet energy scale also show a correlation. It is not surprising to see correlated effects of systematic variations, which are all related to the same object.

In the cut-based analysis, there is a strong negative correlation between the jet energy resolution and the Z normalization in VBF. The reason for this might be event migration into the VBF region when applying this variation.

The multivariate analysis shows a much stronger correlation between the Z normalization and the fake uncertainty in the boosted category compared to the cut-based analysis. This might be explained by the shape difference between the m_{MMC} used for the CBA and the BDT output score for the MVA. The ratio of Z and fake in the output score in the Boosted category is almost constant over the full range, see Fig. 6.18. A change in the fake contribution can be compensated by the Z normalization directly. In contrast to this, in the cut-based analysis a variation of the fake contribution can be constrained by the sidebands of the m_{MMC} distribution.

Figures 7.3 and 7.4 show the comparison of pre fit and post fit systematic uncertainties. This kind of plot is commonly referred to as *pull plot*. The points are all aligned on the central line, because Asimov data was used for this fit. The pull plot for the cut-based analysis shows several constrained nuisance parameters, for example the jet energy resolution, jet energy scale related systematic variations and pile-up reweighting. The reason for these constraints is not completely understood and is beyond the scope of this thesis. The pull plot of the multivariate analysis shows similar constraints. The constraints from the CBA are also present in the MVA. In contrast to the CBA, the MVA shows slight constraints for $E_{\mathrm{T}}^{\mathrm{miss}}$ related systematic variations.

Figures 7.5 show the ranking of the systematic uncertainties with the highest impact on the signal strength. The highest ranked uncertainty is JES flavor composition. Even though the nuisance parameter is constraint, it has the strongest effect on μ. This systematic uncertainty was not used in the fit to assess the BDT

[1] If two variables are statistically independent, it means that their joint probability density function factorizes into two independent probability density functions [33]. There is no statistical correlation between these two variables in that case. If, for example, both variations increase the trigger efficiency, the effect of both variations will be similar. Their effects can therefore be linearly correlated.

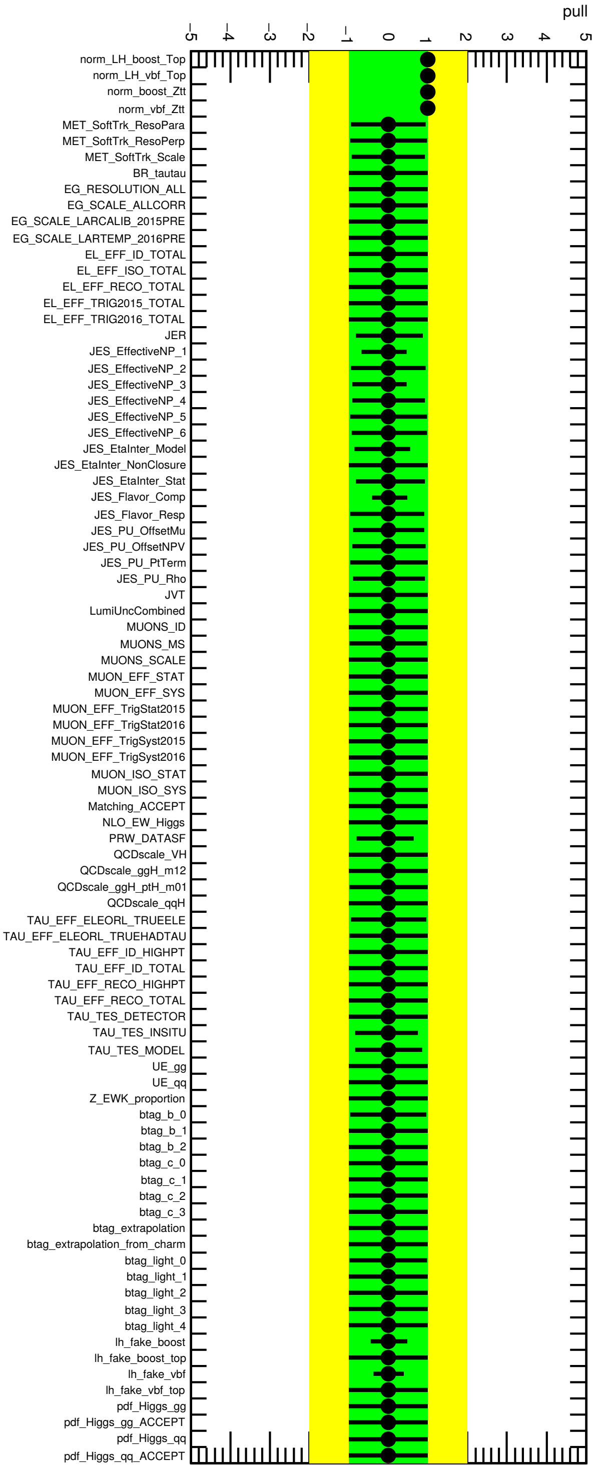

Figure 7.3: Pull plot of the cut-based analysis. The green band shows the pre-fit 1σ systematic variations. The error bar of the black dots indicate the post fit uncertainty.

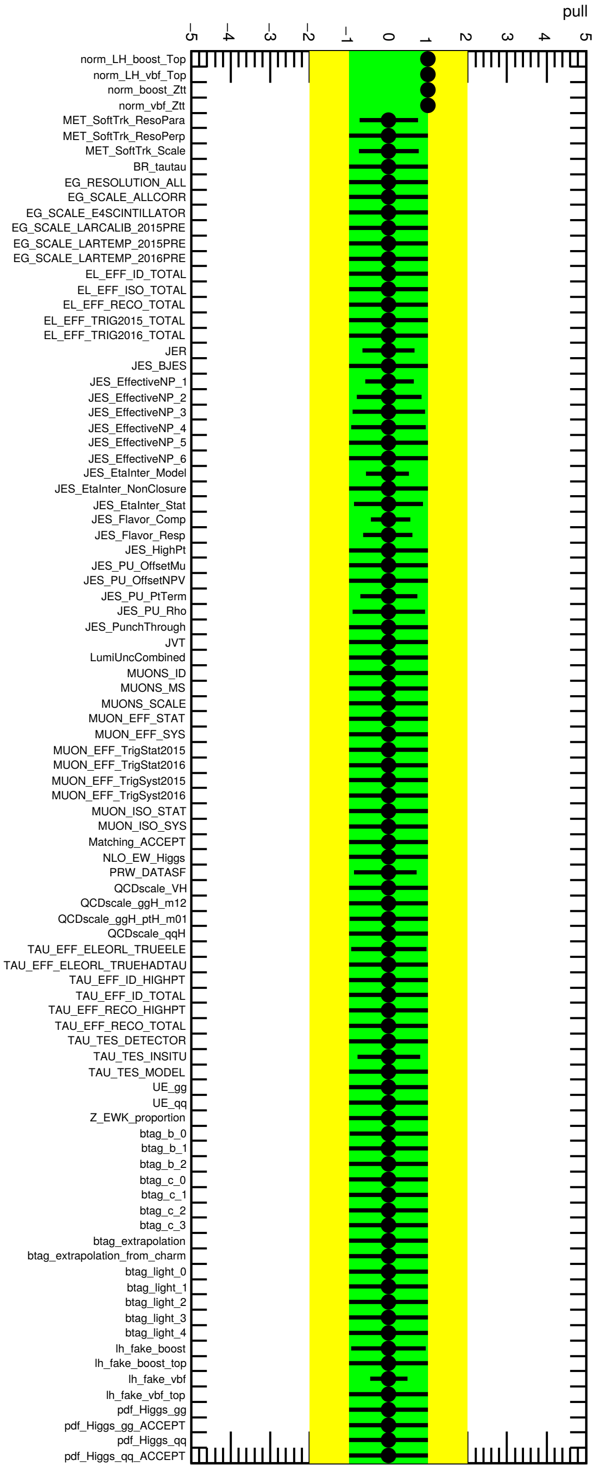

Figure 7.4: Pull plot of the multivariate analysis. The green band shows the pre-fit 1σ systematic variations. The error bar of the black dots indicate the post fit uncertainty.

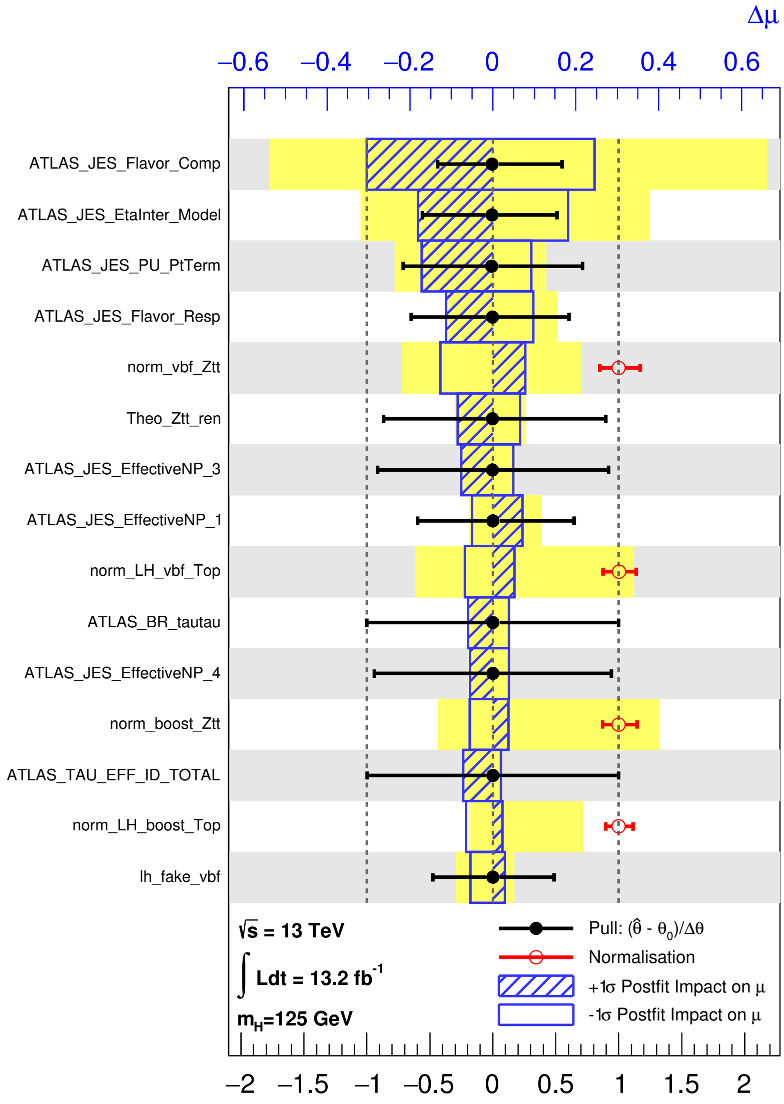

Figure 7.5: Nuisance parameter ranking plot for the multivariate analysis. The yellow band shows the pre-fit uncertainty, the blue boxes show the post fit uncertainty. The plot is limited to the 15 nuisance parameter which the strongest impact on μ.

Table 7.1: Summary of the sensitivity of the cut-based analysis and the multivariate analysis.

	CBA	MVA
Significance Z	1.25	1.63
Uncertainty on μ	± 0.80	± 0.66

performance. The optimization described in Sec. 6.6 could be improved by including this systematic uncertainty in the fit with the reduced set of systematic variations.

Table 7.1 shows the expected significances for both analysis methods derived from the Asmiov fit. The expected significance of the multivariate analysis is about 30% higher than the expected significance of the cut-based analysis.

The multivariate analysis is able to outperform the cut-based analysis in terms of sensitivity. The reason for this can be understood by inspecting the top ranked input variables of the boosted decision trees. Two very important variables are $\ell\,\eta$ centering and $E_{\mathrm{T}}^{\mathrm{miss}}\phi$ centrality, which capture the event topology and have a large discriminating power. The event selection in the cut-based analysis uses the angular information of $\Delta R^{\mathrm{lep\,had}}$, $\Delta\eta^{\mathrm{lep\,had}}$ and $\Delta\eta^{jj}$, whose discriminating powers are inferior to the selection criteria of the multivariate analysis. Furthermore, the multivariate analysis is able to utilize correlations between the input variables, see Figures 6.6 and 6.7, which contributes to the power of the multivariate analysis.

CHAPTER 8

Conclusion

This thesis described a multivariate analysis to search for Higgs boson decays with the two τ leptons in the final state. The analysis was limited to the $H \rightarrow \tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ decay channel. Filtered Monte Carlo samples have been generated, to increase the training statistics in the Boosted category by a factor of approximately ten. The sensitivity of the multivariate analysis has been enhanced by optimizing the BDT training parameters and configuration. The boosted decision trees could be validated by inspecting their behaviour in the control regions. As a comparison a cut-based analysis has also been presented. A likelihood fit has been performed for both analysis methods. The expected significance of the cut-based analysis has been determined to

$$Z_{\mathrm{CBA}} = 1.25. \tag{8.1}$$

The expected significance of the multivariate analysis could be increased to

$$Z_{\mathrm{MVA}} = 1.62, \tag{8.2}$$

therefore the presented multivariate analysis shows a 30% improvement in expected significance over the presented cut-based analysis. Although intensive studies for the multivariate analysis have been carried out, there is still room for improvement in the future.

The most pressing aspect is that this thesis used a partial dataset with an integrated luminosity of $L_{\mathrm{int}} = 13.2\,\mathrm{fb}^{-1}$. During the writing of this thesis the full dataset from 2015 and 2016 has become available for this analysis. With an integrated luminosity of $L_{\mathrm{int}} = 36.1\,\mathrm{fb}^{-1}$, the full dataset offers almost three times as many data.

Another aspect which could be improved in the future is the choices of BDT training parameter configurations in a grid scan. The spacing of the grid scans for the multivariate analysis optimization was rather coarse. In the future a two step procedure could be employed, which adds a fine grid spacing to regions of interest identified by a coarser scan.

The nuisance parameter rankings show the dependence of the signal strength on a certain systematic uncertainty. The ranking shows a large dependence of the JES flavor composition. During the multivariate analysis optimization the performance has been assessed with a likelihood fit, which did not include the JES flavor composition uncertainties. A revised optimization procedure could also take this uncertainty into account and find a BDT more *aware* of this systematic uncertainty.

During the multivariate analysis optimization it became apparent that the small size of the signal event set has a negative effect on the training and thus on the performance of the boosted decision trees. This problem has been mitigated with a privately produced filtered Monte Carlo production for the Boosted category. At this time it would most likely also be beneficial to increase the sample set size for background and VBF signal events. The full dataset from 2015 and 2016 mentioned above comes also with a new production of background and signal Monte Carlo. For background the production consists of filtered Monte Carlo events with increased statistics for the Z boson samples. A dedicated generator filter for VBF signal Monte Carlo can be used to increase the signal training statistics in the VBF category. It can therefore be assumed that the multivariate analysis can benefit from the new samples and might improve further in the future.

The boosted decision trees themselves offer further potential improvements. The software library used in this thesis defines more parameters to control the boosted decision tree training than were used for the multivariate analysis optimization. One approach can be to check if these parameters lead to improved training. Since the last update the this software package was in 2013, another alternative would be to use the Python package *scikit-learn*, which has become very popular in the machine learning community.

The analysis of $H \to \tau\tau$ is not limited to cross section and signal strength measurements. In the future differential cross section measurements or Higgs boson mass measurements with a two-dimensional fit will be pursued. For these kinds of analyses it might be desirable to have a boosted decision tree which does not use m_{MMC} as one of its input variables. Since m_{MMC} is one of the most important variables to discriminate against Z background, this will be a challenging task for the multivariate analysis.

> *"The important thing is not to stop questioning.*
> *Curiosity has its own reason for existence."*
>
> — Albert Einstein

APPENDIX A

Control Regions

Figures A.1, A.2 and A.3 show selected distributions in the W, Z and QCD control regions. Similar distributions for the top control region are shown in Fig. 5.4 in Sec. 5.7.

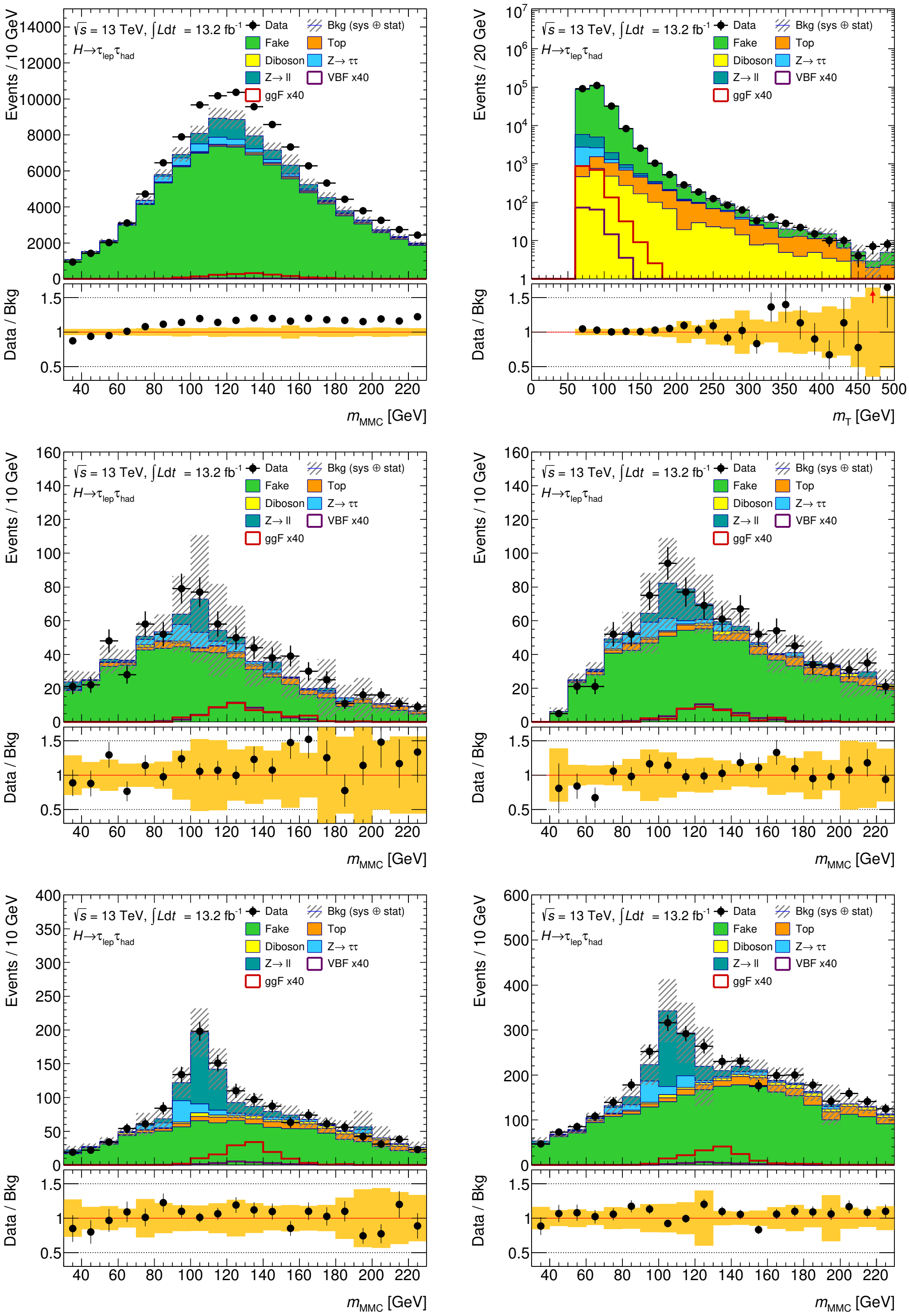

Figure A.1: Selected distributions in the W control region. The error bands include statistical and systematic uncertainties. All plots show the mass m_{MMC}, except the top right plot which shows m_{T}. The top row shows the distributions after applying the preselection cut. The middle row shows the distributions for the VBF categories of the CBA (left) and MVA (right). The bottom row shows the distributions for the Boosted category of the CBA (left) and the MVA (right).

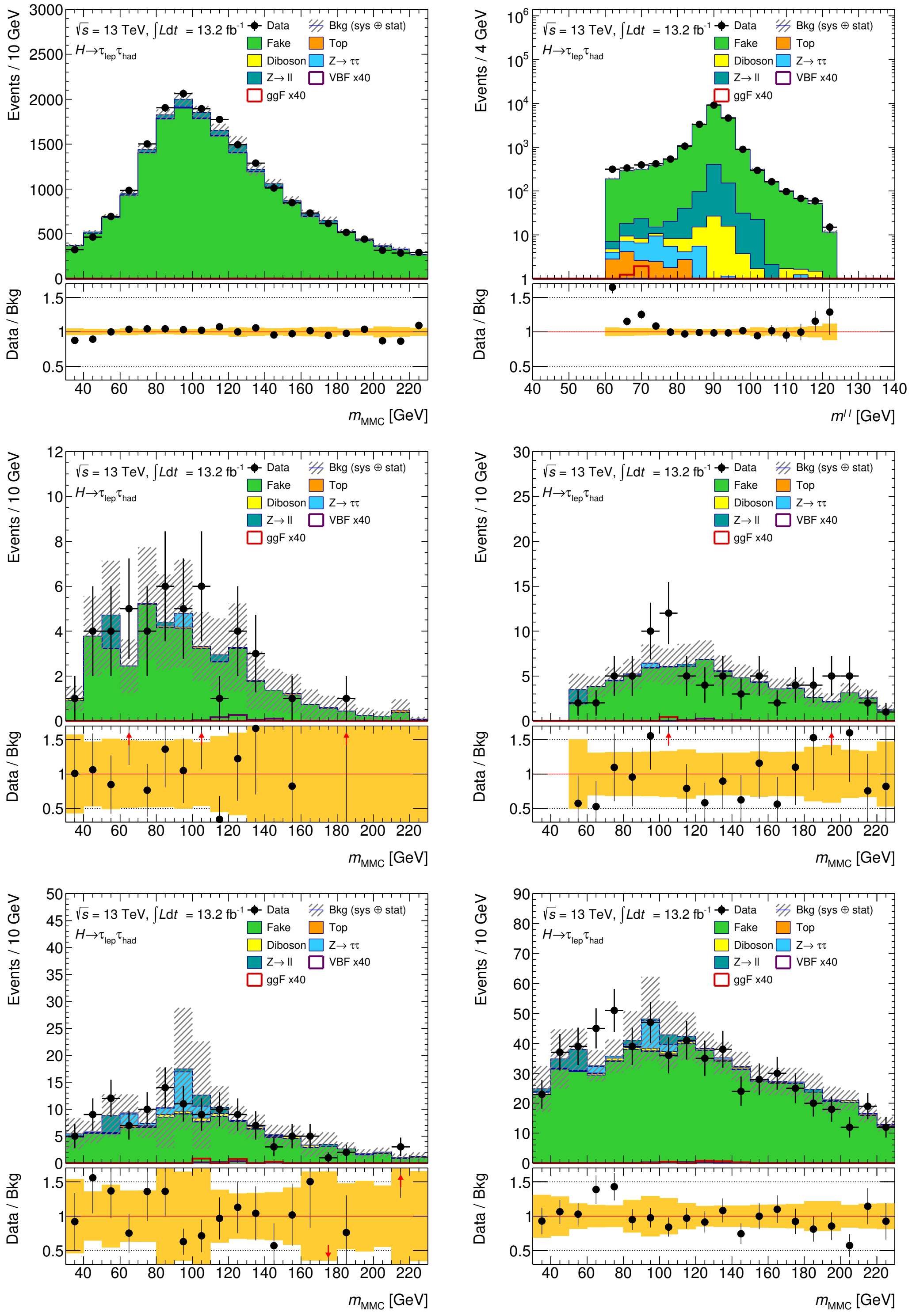

Figure A.2: Selected distributions in the Z control region. The error bands include statistical and systematic uncertainties. All plots show the mass m_{MMC}, except the top right plot which shows the mass of the di-lepton system $m^{\ell\ell}$. The top row shows the distributions after applying the preselection cut. The middle row shows the distributions for the VBF categories of the CBA (left) and MVA (right). The bottom row shows the distributions for the Boosted category of the CBA (left) and the MVA (right).

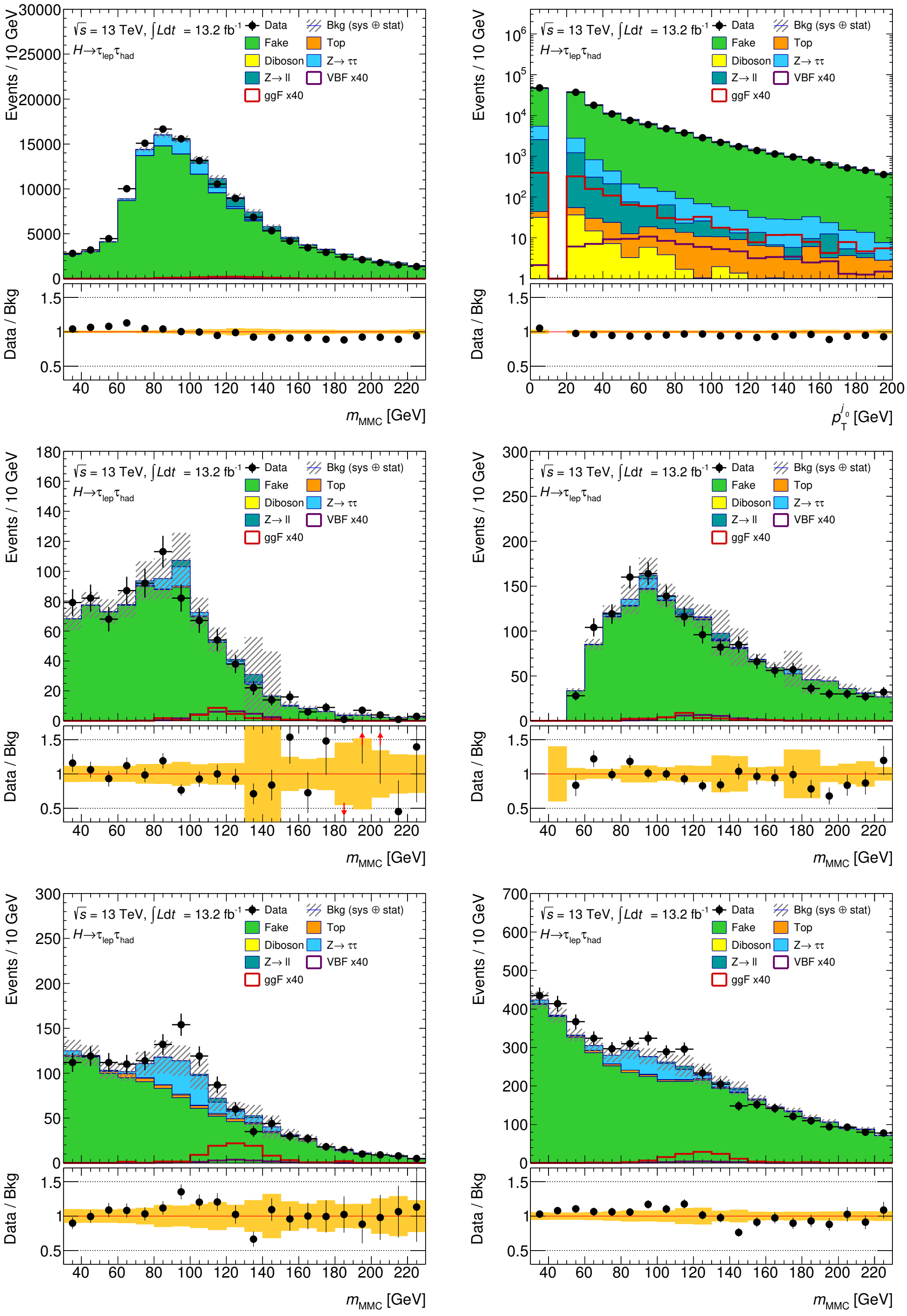

Figure A.3: Selected distributions in the QCD control region. The error bands include statistical and systematic uncertainties. All plots show the mass m_{MMC}, except the top right plot which shows the transverse momentum of the leading jet $p_{\mathrm{T}}^{j_0}$, where entries with $p_{\mathrm{T}}^{j_0} = 0\,\mathrm{GeV}$ indicate that no jet with $p_{\mathrm{T}}^{j_0} \geq 20\,\mathrm{GeV}$ was present in the event. The top row shows the distributions after applying the preselection cut. The middle row shows the distributions for the VBF categories of the CBA (left) and MVA (right). The bottom row shows the distributions for the Boosted category of the CBA (left) and the MVA (right).

APPENDIX B

Variable Ranking

Section 6.3 describes a parameter scan using 47 input variables. Tables B.1 and B.2 show the importance ranking of the input variables derived from this parameter scan for the VBF and Boosted category, respectively. The final input variables have been selected based on the combination of the results from Run 1 and this variable ranking.

Table B.1: Full variable ranking derived from an averaged grid scan for VBF. Variables, which are annotated with ●, are used for the main analysis in this thesis.

Variable	Importance	Variable	Importance
m^{jj}	● 0.0946	h_8	0.0115
$\ell\,\eta$ centrality	● 0.0890	ϕ^{j_1}	0.0110
m_{MMC}	● 0.0866	η^{j_0}	0.0104
$\Delta\eta^{jj}$	● 0.0525	ϕ^{j_0}	0.0102
$E_{\mathrm{T}}^{\mathrm{miss}}\phi$ centrality	● 0.0509	$p_{\mathrm{T}}^{j_0}$	0.0100
$\Delta\eta^{\mathrm{lep\,had}}$	0.0492	ϕ^{j_0}	0.0097
$\Delta R^{\mathrm{lep\,had}}$	● 0.0447	ϕ^{lep}	0.0096
$p_{\mathrm{T}}^{\mathrm{total}}$	● 0.0421	$\lvert\boldsymbol{p}_{\mathrm{T}}^{\mathrm{lep}}+\boldsymbol{p}_{\mathrm{T}}^{\mathrm{had}}\rvert$	0.0096
$m_{\mathrm{vis}}^{\mathrm{lep\,had}}$	0.0366	h_2	0.0094
n^{jets}	0.0311	h_6	0.0094
m_{T}	● 0.0266	$\phi E_{\mathrm{T}}^{\mathrm{miss}}$	0.0093
$E_{\mathrm{T}}^{\mathrm{miss}}$	0.0223	$\Delta p_{\mathrm{T}}^{\mathrm{lep\,had}}$	0.0093
η^{τ}	0.0206	scalar $\sum p_{\mathrm{T}}$	0.0092
p_{T}^{τ}	0.0188	h_7	0.0090
$\Delta\phi^{\mathrm{lep\,had}}$	0.0160	$n^{\mathrm{electrons}}$	0.0090
h_3	0.0155	h_1	0.0089
$p_{\mathrm{T}}^{\mathrm{lep}}/p_{\mathrm{T}}^{\mathrm{had}}$	0.0150	h_5	0.0087
$\eta^{j_0}\cdot\eta^{j_1}$	● 0.0149	$p_{\mathrm{T}}^{\mathrm{lep}}$	0.0081
p_{T}^{H}	0.0140	$p_{\mathrm{T}}^{\mathrm{lep}}+p_{\mathrm{T}}^{\mathrm{had}}$	0.0074
η^{j_1}	0.0137	n^{muons}	0.0065
h_4	0.0127	x_0^{collin}	0.0041
η^{lep}	0.0123	x_1^{collin}	0.0030
$\sum_{\mathrm{jets}} p_{\mathrm{T}}$	0.0121	tight lepton id.	0.0009
$p_{\mathrm{T}}^{j_1}$	0.0120		

Table B.2: Full variable ranking derived from an averaged grid scan for Boosted. Variables, which are annotated with $\bullet$, are used for the main analysis in this thesis.

Variable	Importance
m_{MMC}	$\bullet$ 0.2256
$m_{\mathrm{vis}}^{\mathrm{lep\,had}}$	0.1084
$\Delta R^{\mathrm{lep\,had}}$	$\bullet$ 0.0898
$E_{\mathrm{T}}^{\mathrm{miss}}\phi$ centrality	$\bullet$ 0.0768
$p_{\mathrm{T}}^{\mathrm{lep}}/p_{\mathrm{T}}^{\mathrm{had}}$	$\bullet$ 0.0402
$\Delta\phi^{\mathrm{lep\,had}}$	0.0382
$\Delta\eta^{\mathrm{lep\,had}}$	0.0333
m_{T}	$\bullet$ 0.0283
$E_{\mathrm{T}}^{\mathrm{miss}}$	0.0266
p_{T}^{τ}	0.0188
$\Delta p_{\mathrm{T}}^{\mathrm{lep\,had}}$	0.0165
η^{lep}	0.0159
h_3	0.0152
scalar $\sum p_{\mathrm{T}}$	$\bullet$ 0.0142
η^{τ}	0.0140
$\lvert\boldsymbol{p}_{\mathrm{T}}^{\mathrm{lep}} + \boldsymbol{p}_{\mathrm{T}}^{\mathrm{had}}\rvert$	0.0127
h_2	0.0121
$p_{\mathrm{T}}^{\mathrm{total}}$	0.0120
η^{j_0}	0.0112
n^{jets}	0.0102
p_{T}^{H}	0.0100
$p_{\mathrm{T}}^{\mathrm{lep}}$	0.0099
$p_{\mathrm{T}}^{j_0}$	0.0082
$\eta^{j_0}\cdot\eta^{j_1}$	0.0080
h_1	0.0078
h_4	0.0076
$\sum_{\mathrm{jets}} p_{\mathrm{T}}$	0.0072
h_8	0.0072
ϕ^{j_0}	0.0072
η^{j_1}	0.0071
ϕ^{j_0}	0.0070
$n^{\mathrm{electrons}}$	0.0069
$p_{\mathrm{T}}^{j_1}$	0.0069
$p_{\mathrm{T}}^{\mathrm{lep}} + p_{\mathrm{T}}^{\mathrm{had}}$	0.0069
ϕ^{lep}	0.0068
$\phi^{E_{\mathrm{T}}^{\mathrm{miss}}}$	0.0067
h_7	0.0065
$\Delta\eta^{jj}$	0.0064
h_5	0.0063
ϕ^{j_1}	0.0061
$\ell\,\eta$ centrality	0.0061
n^{muons}	0.0057
h_6	0.0057
m^{jj}	0.0051
x_0^{collin}	0.0041
x_1^{collin}	0.0033
tight lepton id.	0.0008

Bibliography

[1] The ATLAS Collaboration, G. Aad et al., *The ATLAS Experiment at the CERN Large Hadron Collider*, Journal of Instrumentation **3** (2008) S08003.

[2] The CMS Collaboration, G. L. Bayatian et al., *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software.* Technical Design Report CMS. CERN, Geneva, 2006. `https://cds.cern.ch/record/922757`.

[3] The ATLAS Collaboration, G. Aad et al., *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Physics Letters B **716** (2012) 1 – 29.

[4] The CMS Collaboration, S. Chatrchyan et al., *Observation of a new boson at a mass of* 125 GeV *with the CMS experiment at the LHC*, Physics Letters B **716** (2012) 30 – 61.

[5] O. S. Brüning et al., *LHC Design Report.* CERN, Geneva, 2004. `https://cds.cern.ch/record/782076`.

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

[7] P. Higgs, *Broken symmetries, massless particles and gauge fields*, Physics Letters **12** (1964) 132 – 133.

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

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

[10] F. Halzen and A. D. Martin, *Quarks and Leptons: An Introduction Course in Modern Particle Physics.* John Wiley & Sons, Inc., 1st ed., 1984.

[11] G. Arnison et al., *Experimental observation of isolated large transverse energy electrons with associated missing energy at* $\sqrt{s} = 540\,\mathrm{GeV}$, Physics Letters B **122** (1983) 103 – 116.

[12] G. Arnison et al., *Experimental observation of lepton pairs of invariant mass around* 95 GeV/c^2 *at the CERN SPS collider*, Physics Letters B **126** (1983) 398 – 410.

[13] The ATLAS and CMS Collaborations, *Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at* $\sqrt{s} = 7$ *and* 8 TeV, Journal of High Energy Physics **2016** (2016) 45.

[14] The ATLAS Collaboration, G. Aad et al., *Evidence for the Higgs-boson Yukawa coupling to tau leptons with the ATLAS detector*, Journal of High Energy Physics **2015** (2015) 117.

[15] N. Ruthmann and K. Jakobs, *Search for Standard Model* $H \to \tau^+\tau^-$ *Decays in the Lepton-Hadron Final State in Proton-Proton Collisions with the ATLAS Detector at the LHC.* PhD thesis, Freiburg U., Oct, 2014. `https://cds.cern.ch/record/1984325`. Presented 18 Dec 2014.

[16] D. Griffiths, *Introduction to Elementary Particles.* Wiley-VCH, 2nd ed., 2008.

[17] Particle Data Group Collaboration, K. A. Olive et al., *Review of Particle Physics*, Chin. Phys. **C38** (2014) 090001.

[18] T. Plehn, *Lectures on LHC Physics.* Lecture Notes in Physics. Springer International Publishing, 2014.

[19] O. Behnke, K. Kröninger, G. Schott, and T. Schörner-Sadenius, *Data Analysis in High Energy Physics: A Practical Guide to Statistical Methods.* Wiley, 2013.

[20] W. Stirling, *private communication*,.

[21] D. de Florian et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, 2016.

[22] M. Capeans et al., *ATLAS Insertable B-Layer Technical Design Report*, Tech. Rep. CERN-LHCC-2010-013. ATLAS-TDR-19, Sep, 2010. `https://cds.cern.ch/record/1291633`.

[23] A. L. Rosa, *The ATLAS Insertable B-Layer: from construction to operation*, Journal of Instrumentation **11** (2016) C12036.

[24] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-*k_t *jet clustering algorithm*, Journal of High Energy Physics **2008** (2008) 063.

[25] *Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC*, Tech. Rep. ATL-PHYS-PUB-2015-045, CERN, Geneva, Nov, 2015. `https://cds.cern.ch/record/2064383`.

[26] A. Elagin, P. Murat, A. Pranko, and A. Safonov, *A new mass reconstruction technique for resonances decaying to* $\tau\tau$, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **654** (2010) 481 – 489.

[27] L. Spiller, *Modification of Fox-Wolfram moments for hadron colliders*, Journal of High Energy Physics **2016** (2016) 27.

[28] C. Bishop, *Pattern Recognition and Machine Learning.* Information Science and Statistics. Springer, 2006.

[29] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition.* Springer Series in Statistics. Springer New York, 2009.

[30] A. Hoecker et al., *TMVA - Toolkit for Multivariate Data Analysis*, `arXiv:physics/0703039`.

[31] Y. Freund and R. E. Schapire, *A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting*, Journal of Computer and System Sciences **55** (1997) 119 – 139.

[32] The ATLAS Collaboration, M. Aaboud et al., *Luminosity determination in pp collisions at* $\sqrt{s} = 8\,\mathrm{TeV}$ *using the ATLAS detector at the LHC*, The European Physical Journal C **76** (2016) 653.

[33] G. Cowan, *Statistical Data Analysis.* Clarendon Press, 1998.