Masterarbeit

Optimizing the measurement of the signal strength for Higgs-boson production in the $H \to \tau\tau$ decay using multivariate techniques at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector

Benjamin Rottler

Fakultät für Mathematik und Physik
Albert-Ludwigs-Universität Freiburg

Optimizing the measurement of the signal strength for Higgs-boson production in the $H \to \tau\tau$ decay using multivariate techniques at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector

Masterarbeit

vorgelegt von

Benjamin Rottler

20. Oktober 2017

Betreuer

Prof. Dr. Markus Schumacher

Physikalisches Institut
Albert-Ludwigs-Universität
Freiburg im Breisgau

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

.....
Benjamin Rottler

Abstract

The Higgs boson was first observed in July 2012 by the ATLAS and CMS experiments based at CERN [1, 2]. In a combined measurement of the two collaborations the mass was determined to $m_H = (125.09 \pm 0.21(\text{stat.}) \pm 0.11(\text{sys.}))\,\text{GeV}$ [3]. Further measurements confirmed the consistency with Standard-Model predictions for the Higgs boson. The $H \to \tau\tau$ decay channel is the most sensitive decay channel to probe the Yukawa couplings of the Higgs boson. This makes it an interesting channel to analyze during the second data-taking period of the LHC which started in 2015. In this thesis a multivariate approach based on boosted decision trees is developed to increase the sensitivity with respect to the cut-based analysis of the $H \to \tau^+\tau^- \to \ell^+\ell^- 4\nu$ decay channel for a combined 2015 and 2016 dataset corresponding to an integrated luminosity of $\mathcal{L} = 36.1\,\text{fb}^{-1}$ recorded with the ATLAS detector in proton–proton collisions at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$. The BDT hyperparameters and observables used as input variables are optimized in a k-fold cross-validation approach. The expected sensitivity of the cut-based analysis is 0.83σ. Using the multivariate approach it was possible to increase the sensitivity by 63 % to 1.35σ. Additionally, the expected uncertainties on the measurement of the signal strength $\mu = \sigma_{\text{obs}}/\sigma_{\text{SM}}$ are reduced from 1 ± 1.27 in the cut-based analysis to 1 ± 0.68 in the multivariate analysis.

Zusammenfassung

Das Higgs-Boson wurde erstmals im Juli 2012 durch das ATLAS und CMS Experiment am CERN beobachtet [1, 2]. Das kombinierte Ergebnis der ATLAS und CMS Kollaborationen für die Masse des Higgs-Bosons ist $m_H = (125.09 \pm 0.21(\text{stat.}) \pm 0.11(\text{sys.}))\,\text{GeV}$ [3]. Weitere Messungen konnten die Übereinstimmung mit dem durch das Standard Model vorhergesagten Higgs-Bosons bestätigen. Der sensitivste Zerfallskanal für die Messung der Yukawa-Kopplung des Higgs-Bosons ist der $H \to \tau\tau$ Zerfallskanal. Deshalb ist die Analyse dieses Zerfallskanals eine wichtige Aufgabe des zweite Datennahmeperiode am LHC, welche im Jahre 2015 gestartet wurde. In dieser Masterarbeit wird eine multivariate Methode basierend auf verstärkten Entscheidungsbäumen (BDTs) entwickelt, welche die Sensitivität auf das Signal im Vergleich zur Schnitt-basierten Analyse des $H \to \tau^+\tau^- \to \ell^+\ell^- 4\nu$ Zerfalls mit dem kombinierten Datensatz aus den Jahren 2015 und 2016 mit einer integrierten Luminosität von $\mathcal{L} = 36.1\,\text{fb}^{-1}$, welcher mit dem ATLAS Detektor bei einer Schwerpunktsenergie von $\sqrt{s} = 13\,\text{TeV}$ in Proton–Proton Kollisionen aufgenommen wurde, verbessert. Die Hyperparameter und Observablen, die für die BDTs benutzt werden, werden mit einer k-fachen Kreuzvalidierung optimiert. Die erwartete Sensitivität der Schnitt-basierten Analyse von 0.83σ konnte mit dem multivariaten Verfahren um 63 % auf 1.35σ erhöht werden. Des Weiteren wird mit der multivariaten Methode die erwartete Unsicherheit auf die Messung der Signalstärke $\mu = \sigma_{\text{obs}}/\sigma_{\text{SM}}$ von ± 1.27 auf ± 0.68 reduziert.

Contents

1. Introduction

In elementary particle physics the fundamental constituents of nature and their interactions are investigated. In the 1960s and 1970s the Standard Model (SM) of particle physics was developed to provide an accurate description of all elementary particles which are known today and their fundamental interactions. The particles can be classified based on their spin: *fermions* have a half-integer spin and *bosons* have an integer spin. Three out of the four fundamental interactions, the electromagnetic, strong, and weak interaction, are incorporated in the SM as relativistic quantum field theories. No formulation of gravity as a relativistic quantum field theory is found yet. However, the influence of gravity at the energy scales considered in particle physics is minor and therefore can be neglected.
With the help of the Standard Model it was possible to predict several particles, which were later found in nature, for example the massive $W^{\pm}$- and Z^0-bosons which were first observed in 1983 [4–7]. In earlier models, which attempted to describe elementary particles in a quantum field theory, no mass terms for fermions and bosons were included, which stands in contrast to measurements of fermion and boson masses. For example, the $W^{\pm}$- and Z^0-bosons have a mass of $m_{W^{\pm}} = 80.4\,\text{GeV}$ and $m_{Z^0} = 91.2\,\text{GeV}$, respectively [8]. This conflict was solved by the introduction of the Englert–Brout–Higgs–Guralnik–Hagen–Kibble mechanism [9–14] (short: Higgs mechanism) which uses the principle of spontaneous symmetry breaking. A new scalar field, the Higgs field, is introduced. The interaction of fermions and bosons with the vacuum expectation value of the Higgs field leads to the required mass terms of the particles. The particle associated with the Higgs field is the *Higgs boson*. During the last decades great effort was put into the search for the Higgs boson at different collider experiments, until it was found with the ATLAS[1] and CMS[2] experiments at CERN[3] in 2012 [1, 2]. The mass of the newly found particle was determined to be $m_H = 125.09 \pm 0.21(\text{stat.}) \pm 0.11(\text{sys.})\,\text{GeV}$ in a combined measurement of the ATLAS and CMS collaboration [3]. After the Higgs boson was found at CERN, Peter Higgs and Francois Englert were awarded the Nobel Prize in 2013 for the formulation of the underlying theory.
The $H \to \tau\tau$ decay mode is an important decay channel of the Higgs boson, since it it provides the most sensitive measurement of the Yukawa couplings of the Higgs boson. Additional, the $H \to \tau\tau$ decay channel can be used to probe a potential CP mixing and lepton-flavour violating decays of the Higgs boson. During the first data-taking period at the LHC in 2011 and 2012 evidence for this decay channel was found at the ATLAS experiment at the level of 4.5σ [15]. The signal strength

[1] **A** **T**oroidal **L**HC **A**pparatu**S**
[2] **C**ompact **M**uon **S**olenoid
[3] **C**onseil **E**uropéen pour la **R**echerche **N**ucléaire

$\mu = \sigma_{\text{obs}}/\sigma_{\text{SM}}$ was determined to be $\mu = 1.43^{+0.43}_{-0.37}$. Due to limited statistic the observation where a sensitivity of 5σ is required could not be made.
In 2015 the second data taking period started at the LHC with an increased center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ and integrated luminosity. This allows for a more precise measurement of the signal strength of the $H \to \tau\tau$ channel and measurements in the individual production modes. However, due to the new conditions the analysis strategy needs to be reoptimized. In this thesis the $H \to \tau^+\tau^- \to \ell^+\ell^- 4\nu$ decay channel is considered, where both leptons decay leptonically, with the combined 2015 and 2016 dataset of the second data-taking period corresponding to an integrated luminosity of $\mathcal{L} = 36.1\,\text{fb}^{-1}$ at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$. A method is developed to increase the sensitivity in this decay channel with the help of multivariate techniques. More precisely, the usage of boosted decision trees (BDTs), a machine-learning algorithm, is investigated.
This thesis is structured as follows. First the Standard Model is introduced and an overview of the current status of Higgs-boson property measurements is given in Chapter 2. In Chapter 3 the LHC and ATLAS experiment are described. The signal and background processes considered in the analysis of $H \to \tau^+\tau^- \to \ell^+\ell^- 4\nu$ are discussed in Chapter 4, followed by an overview of the reconstruction of physical objects in Chapter 5. The event selection is described in Chapter 6 and the strategy to estimate specific backgrounds is presented in Chapter 7. An overview of the theoretical aspects of BDTs is given in Chapter 8 and their application to the analysis of the $H \to \tau^+\tau^- \to \ell^+\ell^- 4\nu$ process is discussed in Chapter 9. In Chapter 10 the systematic uncertainties are presented. The thesis closes with a description of the statistical procedure and a discussion of the results in Chapter 11, followed by a summary of the analysis and an outlook for further studies in Chapter 12.

2. Theory

This chapter introduces the theoretical background on which this analysis is based on. First, the Standard Model (SM) of particle physics, which describes the fundamental particles and their interactions, is discussed. Then, an overview of the phenomenology of hadron scattering is given. Afterwards the Higgs-boson production and decay is discussed in detail and past measurements of the Higgs boson are presented.

2.1. The Standard Model of Particle Physics

The Standard Model of particle physics was developed during the 1960s and 1970s and describes elementary particles and their fundamental interactions to great precision. It is a relativistic quantum field theory based on the principle of local gauge symmetries. Several particles like the $W^{\pm}$- and Z^0- bosons and the top-quark were predicted by the SM and later discovered in nature [4–7, 16, 17]. Out of the four fundamental forces it describes only three, since it was not yet successful to give a formulation of gravity as a quantum field theory. But because gravity has only a minor effect on particles at the energy scales accessible at current collider experiments it can be neglected.
The contents of this section are based on [18–20] if not noted otherwise.

2.1.1. Elementary Particles

The elementary particles of the SM can be divided into fermions which carry half integer spin and bosons with a full integer spin. Fermions can further be split into leptons which only interact via the electroweak force and quarks which interact via both the electroweak and strong force. Empirical evidence shows that there are three generations of both leptons and quarks. This cannot be predicted by the SM, it was rather used as input. The three generations are copies of each other which are identical expect for the flavour and mass of the particles. They are sorted in ascending order by the masses of the particles. There exist two elementary particles for each generation.
In the case of leptons there is the electron (e), muon (μ), and tau lepton (τ) which all have a electric charge[1] of $Q = -1$. Each lepton is associated with a neutral particle called the neutrino (ν_e, ν_μ, ν_τ).
For quarks one particle of each flavour has a electric charge of $Q = \frac{2}{3}$. These particles are called the up-, charm-, and top-quark. The other quarks are the

[1] The electric charge is given in units of the fundamental electric charge $q_e = 1.602 \times 10^{-19}$ C [8].

down-, strange-, and bottom-quark with charge of $Q = -\frac{1}{3}$. An overview of the fermions in the SM with their charges and approximate masses is given in Table 2.1. Each particle has its corresponding anti-particle, which has reversed quantum numbers in properties of charge and magnetic momentum.
The matter surrounding us consists only of fermions from the first generation, namely the up- and down-quark which form protons and neutrons and the electron. All other particles are only accessible via cosmic radiation or accelerator experiments.

Table 2.1.: Overview of the fermions in the Standard Model [8].

	Generation	Flavour		Charge [q_e]	Mass [GeV]
Leptons	1^{st}	e	Electron	-1	$\approx 0.5 \times 10^{-3}$
		ν_e	Electron neutrino	0	$< 2 \times 10^{-9}$
	2^{nd}	μ	Muon	-1	$\approx 106 \times 10^{-3}$
		ν_μ	Muon Neutrino	0	$< 0.19 \times 10^{-3}$
	3^{rd}	τ	τ-lepton	-1	≈ 1.777
		ν_τ	τ-lepton neutrino	0	$< 18 \times 10^{-3}$
Quarks	1^{st}	u	Up	$\frac{2}{3}$	$\approx 2.2 \times 10^{-3}$
		d	Down	$-\frac{1}{3}$	$\approx 4.7 \times 10^{-3}$
	2^{nd}	c	Charm	$\frac{2}{3}$	≈ 1.28
		s	Strange	$-\frac{1}{3}$	$\approx 96 \times 10^{-3}$
	3^{rd}	t	Top	$\frac{2}{3}$	≈ 173
		b	Bottom	$-\frac{1}{3}$	≈ 4.18

The fundamental interactions of the SM are mediated by gauge bosons with spin one. An overview of the gauge bosons in the SM is given in Table 2.2.
The photon (γ) is the mediator of the electromagnetic force. It couples to the electric charge Q of the particles. Due to the fact that the photon is massless, the electromagnetic force has infinite range.
The weak interaction is transmitted via the massive $W^{\pm}/Z^0$-bosons, which couple to the weak isospin I_w of the particles. All fermions have a weak isospin of $I_w = \frac{1}{2}$ and the bosons of $I_w = 1$. In contract to the photon the $W^{\pm}$- and Z^0-bosons are quite massive with a mass of $m_{W^{\pm}} = 80.4\,\text{GeV}$ and $m_{Z^0} = 91.2\,\text{GeV}$, respectively [8]. This leads to the weak coupling at low energies and the low range of the weak interaction. Additionally, the $W^{\pm}/Z^0$-bosons can interact with themselves.
The strong force is mediated by eight gluons (g), which are massless and couple to the color charge. All quarks and gluons have a color charge. Because the gluons have a color charge themselves, self-interaction is possible, which limits the range of the strong force.
The Standard Model contains also one scalar spin-0 particle, the Higgs boson. It has been detected in 2012 with the ATLAS and CMS detectors [1, 2]. The Higgs

Table 2.2.: Overview of the gauge bosons in the Standard Model [8].

Interaction	Gauge boson		Charge [q_e]	Mass [GeV]	Range [m]
Electromagnetic	γ	Photon	0	0	∞
Weak	$W^{\pm}$	$W^{\pm}$ boson	± 1	80.4	$< 10^{-15}$
	Z^0	Z^0 boson	0	91.2	
Strong	g	Gluon (8x)	0	0	$\approx 10^{-15}$

boson has a mass of $m_H = 125\,\text{GeV}$ [3], no electric or color charge, a weak isospin of $I_w = \frac{1}{2}$, and a weak hypercharge of $Y = 1$. The Higgs boson and the current measurements of Higgs-boson properties are discussed in detail in Sections 2.2 and 2.3.

2.1.2. Fundamental interactions

A gauge theory is a field theory where the Lagrangian density is invariant under a group of global or local transformations. In the Standard Model the fundamental interactions are described by local gauge symmetries. The symmetries force a certain structure on the Lagrangian, which reflects in the resulting theory. The correct symmetry group has to be selected based on empirical observations, so that theory predictions are in agreement with measurements.

Quantum Electrodynamics

The electromagnetic interaction is described by Quantum Electrodynamics (QED), which is based on the $U(1)_Q$ symmetry group. It is carried by the photon, which is massless and couples to the electric charge Q. A free fermion with mass m can be described by the Lagrangian density

$$\mathcal{L} = \bar{\psi}\left(i\gamma^\mu \partial_\mu - m\right)\psi\,, \tag{2.1}$$

where the Dirac spinor is denoted as ψ, the gamma matrices as γ^μ, and the partial derivative as $\partial_\mu = \frac{\partial}{\partial x^\mu}$. By applying the Euler–Lagrange equation,

$$\frac{\partial \mathcal{L}}{\partial \bar{\psi}} - \partial^\mu \left(\frac{\partial \mathcal{L}}{\partial\left(\partial^\mu \bar{\psi}\right)} \right) = 0\,, \tag{2.2}$$

the corresponding equation of motion can be obtained,

$$\left(i\gamma^\mu \partial_\mu - m\right)\psi = 0\,. \tag{2.3}$$

Because every gauge theory is required to be invariant under local transformations of the symmetry group, the lagrangian density in Eq. (2.1) has to be invariant under local transformations of the $U(1)_Q$ group, which have the following form:

$$\psi \mapsto \psi' = e^{-iQ\alpha(x)}\psi\,. \tag{2.4}$$

The operator of electric charge is denoted as Q and the local phase depending on time and space as $\alpha(x)$. However, if this transformation is applied to Eq. (2.1) an additional term appears and the gauge invariance is broken,

$$\partial_\mu \psi \mapsto \partial_\mu \psi' = -iQ\alpha(x)\psi + e^{-iQ\alpha(x)}\partial_\mu \psi \,. \tag{2.5}$$

To restore the gauge invariance a new vector field A_μ, which transforms as

$$A_\mu \mapsto A'_\mu = A_\mu + \partial_\mu \alpha(x) \,, \tag{2.6}$$

and the covariant derivative,

$$D_\mu = \partial_\mu + iqA_\mu \,, \tag{2.7}$$

need to be introduced. The gauge-invariant Lagrangian density of QED reads

$$\mathcal{L} = \bar{\psi}\left(i\gamma^\mu D_\mu - m\right)\psi \,. \tag{2.8}$$

The new vector field A_μ couples to fermions with a coupling strength of Q_f and ensure the local gauge invariance. It can be identified with the photon when a kinematic term, which is formed by the field strength tensor

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \,, \tag{2.9}$$

is added to the Lagrangian density,

$$\mathcal{L}_{\text{QED}} = \bar{\psi}\left(i\gamma^\mu D_\mu - m\right)\psi - \frac{1}{4}F^{\mu\nu}F_{\nu\nu} \,. \tag{2.10}$$

If a mass term of the form $-\frac{1}{2}m^2 A^\mu A_\mu$ was introduced, it would break gauge invariance again. Therefore, a massless photon is required in QED, which corresponds with the upper limit of the photon mass of $m_\gamma < 3 \times 10^{-27}\,\text{eV}$ obtained by experimental measurements [21].

Quantum Chromodynamics

The interaction of quarks and gluons via the strong force is described by Quantum Chromodynamics (QCD), which is based on the $SU(3)_C$ symmetry group. The charge associated with the strong interaction is the color charge, which is the equivalent to the electric charge in QED. Experimental measurements show that there are three different color states: red, green, and blue. Those three states can be described by building a vector of three spinor fields, which replaces the single Dirac spinor ψ from QED,

$$\psi = \begin{pmatrix} \psi_{\text{red}} \\ \psi_{\text{green}} \\ \psi_{\text{blue}} \end{pmatrix} \,. \tag{2.11}$$

Under a local $SU(3)_C$ transformation a free quark field $\psi(x)$ transforms like

$$\psi(x) \mapsto \psi'(x) = \exp\left[i\frac{g_s}{2}\sum_{a=1}^{8}\lambda_a \beta_a(x)\right]\psi(x) \,. \tag{2.12}$$

Here, the coupling strength is denoted as α_s, the eight Gell-Mann matrices as λ_a, and the β functions of QCD as $\beta(x)$ [22]. In the following the gauge coupling parameter $g_s = \sqrt{4\pi\alpha_s}$ is used instead of the coupling strength.
The $SU(3)_C$ group is a *non-abelian* group, since its generators do not commute. This results in an additional term in the field strength tensor $G^a_{\mu\nu}$ of the gluon fields G^a_μ $(a = 1, \ldots, 8)$,

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - g_s f^{abc} G^b_\mu G^c_\nu \,, \tag{2.13}$$

with the structure constants f^{abc} of the $SU(3)_C$ group. Because there are eight Gell-Mann matrices, which are the generators of $SU(3)_C$, there are also eight gluon fields defined.
To ensure gauge invariance again a covariant derivative is introduced,

$$D_\mu = \partial_\mu + i g_s \frac{\lambda_a}{2} G^a_\mu \,. \tag{2.14}$$

The Lagrangian density of QCD can then be written as

$$\mathcal{L}_{\text{QED}} = \bar{\psi} \left(i\gamma^\mu D_\mu - m \right) \psi - \frac{1}{4} G^{\mu\nu} G_{\nu\nu} \,. \tag{2.15}$$

The non-abelian structure of $SU(3)_C$ leads to gluon self-interaction. Similarly to photons, the gluons need also to be massless to ensure the gauge invariance, which agrees with experimental observations. Because of the gluon self-interaction the strong force has not unlimited range. At very short distances, the strong force becomes weak, which is also known as *asymptotic freedom* [23, 24]. For interactions and long distances the interaction potential increases for color-charged particles. Therefore, free quarks are not stable but form colorless bound states which are called mesons (quark and anti-quark) and baryons (three quarks or three anti-quarks). This is called *confinement*.
For quarks a mass term is allowed and does not break the symmetry, unlike for gluons. The masses are different for each flavour but do not depend on the color charge.

Electroweak Interaction

The weak interaction is mediated by the charged $W^\pm$-bosons and the neutral Z^0-boson, which couple to the weak isospin, I_w. The exchange of a $W^\pm$-boson is called *charged current*, because it modifies the flavour of fermions and causes a change in the electric charge of $\Delta Q = 1$. In contrast, the exchange of a Z-boson does not change the flavour of quarks, which leads to the name *neutral current*. It was discovered that weak interactions mediated by $W^\pm$-bosons are maximally parity violating, because the bosons couple only to left-handed particles and right-handed anti-particles [25, 26]. This lead to the combination of the electromagnetic and weak interaction by Glashow, Salam, and Weinberg in the so-called electroweak Standard Model [27–29].

Table 2.3.: Overview of singlets and doublets in the electroweak theory and their associated quantum numbers.

	Generations			Quantum numbers			
	1^{st}	2^{nd}	3^{rd}	I_w	I_3	Y	Q $[q_e]$
Leptons	$\begin{pmatrix}\nu_e\\ e\end{pmatrix}_L$	$\begin{pmatrix}\nu_\mu\\ \mu\end{pmatrix}_L$	$\begin{pmatrix}\nu_\tau\\ \tau\end{pmatrix}_L$	$\frac{1}{2}$ $\frac{1}{2}$	$\frac{1}{2}$ $-\frac{1}{2}$	-1 -1	0 -1
	e_R	μ_R	τ_R	0	0	-2	-1
Quarks	$\begin{pmatrix}u\\ d'\end{pmatrix}_L$	$\begin{pmatrix}c\\ s'\end{pmatrix}_L$	$\begin{pmatrix}t\\ b'\end{pmatrix}_L$	$\frac{1}{2}$ $\frac{1}{2}$	$\frac{1}{2}$ $-\frac{1}{2}$	$\frac{1}{3}$ $\frac{1}{3}$	$\frac{2}{3}$ $-\frac{1}{3}$
	u_R	c_R	t_R	0	0	$\frac{4}{3}$	$\frac{2}{3}$
	d_R	s_R	b_R	0	0	$-\frac{2}{3}$	$-\frac{1}{3}$

The electroweak interaction is based on an underlying $SU(2)_{L,I_w} \times U(1)_Y$ symmetry and is able to describe both the electromagnetic and weak interaction. Here, the hypercharge Y was introduced and L denotes the coupling to only left-handed particles. A connection between the electric charge Q, the hypercharge Y, and the third component of the weak isospin I_3 is given by the Gell-Mann–Nishijima formula [30, 31],

$$Q = I_3 + \frac{Y}{2}\,. \tag{2.16}$$

Left-handed fermions are described by $SU(2)_{L,I_w}$ doublets with $I_w = \frac{1}{2}$ and $I_3 = \pm\frac{1}{2}$. Right-handed fermions are assigned to $SU(2)_{L,I_w}$ singlets with $I_w = I_3 = 0$. An overview of the quantum numbers of fermions in the electroweak theory is given in Table 2.3. Right-handed neutrinos are not included, since they do not couple to other particles of the SM. Recent results from neutrino-oscillation experiments which also yielded a Nobel Price in Physics in 2015 show that at least two neutrino masses are not zero [32–37]. However, in this thesis neutrinos are assumed to be massless.

Quarks are described in weak eigenstates (d', s', b') with $I_3 = -\frac{1}{2}$ which are a mixture of their mass eigenstates (u, s, b). The degree of mixing is described by the Cabibbo–Kobayashi–Maskawa (CKM) matrix [38, 39],

$$\begin{pmatrix}d'\\ s'\\ b'\end{pmatrix} = \begin{pmatrix}V_{ud} & V_{us} & V_{ub}\\ V_{cd} & V_{cs} & V_{cb}\\ V_{td} & V_{ts} & V_{tb}\end{pmatrix}\begin{pmatrix}d\\ s\\ b\end{pmatrix}\,. \tag{2.17}$$

The elements $\left|V_{ij}\right|^2$ give the probability of a quark changing its flavour from i to j when interacting with a $W^{\pm}$-boson. Due to the non-vanishing complex phase of the CKM matrix the CP invariance is violated [40].

Left-handed isospin doublets transform under the $SU(2)_{L,I_w}$ symmetry as

$$\psi_L(x) \mapsto \psi'_L(x) = \exp\left[i\frac{g}{2}\sum_{a=1}^{3}\tau_a\alpha_a(x)\right]\psi_L(x)\,, \tag{2.18}$$

with the generators $\tau_a/2$ $(a = 1, 2, 3)$ of the $SU(2)_{L,I_w}$ symmetry which are the 2×2 Pauli matrices, the coupling strength g, and the local phase $\alpha_a(x)$. The left-handed isospin doublets and the right-handed singlets transform have the following transformation behavior under the $U(1)_Y$ symmetry:

$$\psi_L(x) \mapsto \psi'_L(x) = \exp\left[i\frac{g'}{2}Y\beta(x)\right]\psi_L(x)\,, \tag{2.19}$$

$$\psi_R(x) \mapsto \psi'_R(x) = \exp\left[i\frac{g'}{2}Y\beta(x)\right]\psi_R(x)\,, \tag{2.20}$$

where g' is a second coupling constant, Y the generator of the hypercharge, and $\beta(x)$ the local phase.
To preserve gauge invariance three vector fields W^a $(a = 1, 2, 3)$ for $SU(2)_{L,I_w}$ and one gauge field B for $U(1)_Y$ need to be introduced. With the covariant derivatives for left- and right-handed fermion fields,

$$D^L_\mu = \partial_\mu + i\frac{g}{2}\tau_a W^a_\mu + i\frac{g'}{2}YB_\mu\,, \tag{2.21}$$

$$D^R_\mu = \partial_\mu + i\frac{g'}{2}YB_\mu\,, \tag{2.22}$$

the Lagrangian density for the electroweak interaction reads

$$\mathcal{L}_{\mathrm{EW}} = \bar{\psi}_L i\gamma^\mu D^L_\mu \psi_L + \bar{\psi}_R i\gamma^\mu D^R_\mu \psi_R - \frac{1}{4}W^a_{\mu\nu}W^{\mu_\nu,a} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}\,. \tag{2.23}$$

Here, the field strength tensors are defined as

$$W^a_{\mu\nu} = \partial_\mu W^a_\nu - \partial_\nu W^a_\mu - g\epsilon^{abc}W^b_\mu W^c_\nu\,, \tag{2.24}$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu\,, \tag{2.25}$$

with the structure constants ϵ^{abc} of the $SU(2)_{L,I_w}$ group. The third term in Eq. (2.24) enables self-interaction of the vector fields W^a_μ, while the B_μ field can only interact with fermions.
Because the electroweak theory combines the electromagnetic and weak interaction, it should yield the photon field A^μ. However, since B_μ and W^3_μ both couple to neutrinos, they cannot be identified with A^μ. Only a linear combination of those two fields can lead to the photon field. Of course, the linear combination needs to yield the same coupling properites as A^μ, i.e. it needs to couple to right- and left-handed fermions with the same coupling strength but is not allowed to interact with neutrinos. Additionally, it has to be orthogonal to the field of the Z^0-boson. A weak mixim angle θ_w is introduced,

$$\cos(\theta_w) = \frac{g}{\sqrt{g^2 + g'^2}}\,. \tag{2.26}$$

The photon field A^μ and the Z^0-boson field Z^μ can now be constructed as a mixing of the W^3_μ and B_μ field from the $SU(2)_{L,I_w} \times U(1)_Y$ symmetry,

$$\begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} \cos(\theta_w) & -\sin(\theta_w) \\ \sin(\theta_w) & \cos(\theta_w) \end{pmatrix} \begin{pmatrix} W^3_\mu \\ B_\mu \end{pmatrix}\,. \tag{2.27}$$

Furthermore, the coupling strength e of the electromagnetic interaction can be written as a function of the coupling constants g and g' of the $SU(2)_{L,I_w}$ and $U(1)_Y$ transformations,

$$e = \frac{gg'}{\sqrt{g^2 + g'^2}} = g' \cos(\theta_w) = g \sin(\theta_w) \,. \tag{2.28}$$

The charge eigenstates of the $W^{\pm}$-bosons are formed from a superposition of the W^1_μ and W^2_μ fields,

$$W^{\pm}_\mu = \frac{1}{\sqrt{2}} \left(W^1_\mu \mp i W^2_\mu \right) \,. \tag{2.29}$$

In this electroweak theory all gauge bosons and fermions are required to be massless, because any mass term in Eq. (2.23) would lead to symmetry breaking. This is of course not in agreement with the observation of massive fermions and the masses of the $W^{\pm}$- and Z^0-bosons [8]. This conflict of theory and experiment is resolved by the Englert–Brout–Higgs–Guralnik–Hagen–Kibble[2] mechanism. It introduces four new scalar field in the context of spontaneous symmetry breaking and is discussed in the next section.

2.1.3. Spontaneous Symmetry Breaking and Higgs Mechanism

The Higgs-mechanism [9–14] was developed in 1964 in a quest to solve the problem that no massive fermions and gauge bosons are allowed in the theory of electroweak interaction, which clearly contradicts experimental measurements. With the concept of spontaneous symmetry breaking it is possible to include mass terms in the electroweak part of the Standard Model. The mass terms are introduced by spontaneously breaking the symmetry with a state of minimum energy, the so-called *vacuum state*, of a doublet of complex scalar fields with four degrees of freedom, which transforms under the $SU(2)_{L,I_w}$ symmetry. This field is called the *Higgs field*. Its quantum numbers are $Y = 1$ and $I_w = \frac{1}{2}$ and it can be written as [22]

$$\Phi = \begin{pmatrix} \Phi^+ \\ \Phi^0 \end{pmatrix} = \begin{pmatrix} \Phi_1 + i\Phi_3 \\ \Phi_2 + i\Phi_4 \end{pmatrix} \,, \qquad \Phi_i \in \mathbb{R} \,. \tag{2.30}$$

The corresponding Lagrangian density reads

$$\mathcal{L} = \left(\partial_\mu \Phi \right)^\dagger \left(\partial^\mu \Phi \right) - V(\Phi) \,. \tag{2.31}$$

This Lagrangian density also has to be invariant under transformations of the $SU(2)_{L,I_w} \times U(1)_Y$ symmetry, thus the normal derivatives ∂_μ are replaced with the covariant derivative of the electroweak theory given in Eq. (2.21),

$$\mathcal{L} = \left(D_\mu \Phi \right)^\dagger \left(D^\mu \Phi \right) + \mu^2 \Phi^\dagger \Phi - \lambda \left(\Phi^\dagger \Phi \right) - \frac{1}{4} W^a_{\mu\nu} W^{\mu_\nu, a} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} \,. \tag{2.32}$$

[2] For simplicity this mechanism will be referred to as *Higgs mechanism*.

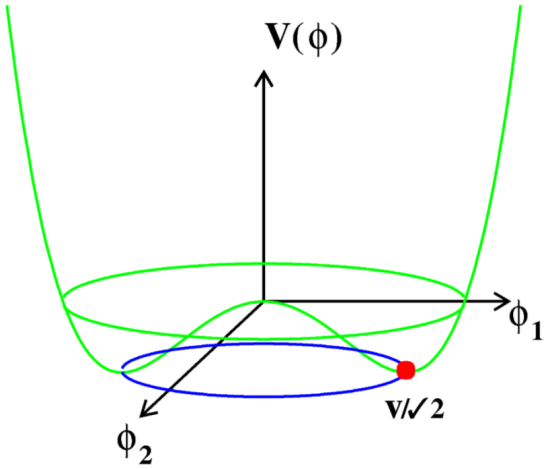


Figure 2.1.: Higgs potential for a scenario of spontaneous symmetry breaking. Only two degrees of freedom are shown [41].

Thus, the most general form of the Higgs potential $V(\Phi)$ which still is invariant under the $SU(2)_{L,I_w} \times U(1)_Y$ symmetry and providing renormalizability is given by

$$-\mu^2 \Phi^\dagger \Phi + \lambda \left(\Phi^\dagger \Phi\right) , \qquad \lambda > 0, \mu^2 > 0 . \tag{2.33}$$

It has a minimum for non-vanishing values of μ, which corresponds to a broken $SU(2)_{L,I_w} \times U(1)_Y$ symmetry. A two-dimensional illustration of the Higgs potential is shown in Fig. 2.1. The minimum of the potential has a radial symmetry. Spontaneous symmetry breaking refers to choosing a specific value for the minimum of the potential.
The Higgs potential can be minimized with respect to $\Phi^\dagger \Phi$ by

$$|\Phi_0|^2 = \frac{v}{\sqrt{2}} , v = \sqrt{\frac{\mu^2}{\lambda}} . \tag{2.34}$$

Here, v is the vacuum expectation value. The usual choice of the ground state is $\Phi_1 = \Phi_2 = \Phi_4 = 0$ and $\Phi_3 = |\Phi_0|$, which leads to

$$\Phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} . \tag{2.35}$$

The ground state has the quantum numbers $Y = 0$ and $I_3 = -\frac{1}{2}$. This breaks the $SU(2)_{L,I_w} \times U(1)_Y$ symmetry spontaneously to $U(1)_Q$. However, the $U(1)_Q$ needs to remain unbroken, therefore Φ_1 and Φ_2 are set to zero, to obtain a neutral ground state.
The vacuum expectation value v can bet set in relation to the Fermi constant G_F [8],

$$v = \left(\sqrt{2} G_F\right)^{-\frac{1}{2}} = 246\,\mathrm{GeV} . \tag{2.36}$$

To describe the potential around the ground state Φ_0, the following parametrization can be made:

$$\Phi(x) = \frac{1}{\sqrt{2}} \exp\left[i \sum_{a=1}^{3} \frac{\tau_a G_a(x)}{v} \right] \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} . \tag{2.37}$$

Four new real scalar fields need to be introduced. The three fields $G_a(x), (a = 1, 2, 3)$ can be associated with the massless scalar Goldstone bosons [42, 43]. They can be eliminated by using a unitary gauge transformation of the form

$$\Phi(x) \mapsto \Phi'(x) = \exp\left[-i \sum_{a=1}^{3} \frac{\tau_a G_a(x)}{v} \right] \Phi(x) . \tag{2.38}$$

The fourth field, $H(x)$, can be interpreted as an excitation of the ground state. This excitation can be associated with a new scalar particle, the Higgs boson.
Using the definitions of the gauge boson fields from Eqs. (2.27) and (2.29), the Lagrangian density can be expanded as,

$$\begin{aligned} \mathcal{L}_{\text{Higgs}} =& \frac{1}{2}(\partial_\mu H)(\partial^\mu H) - \lambda v^2 H^2 - \lambda v H^3 - \frac{1}{4}\lambda H^4 \\ &+ \left(\frac{1}{2} v g\right)^2 W_+^\mu W_\mu^- + \frac{1}{2}\left(\frac{v g}{2\cos(\theta_w)}\right)^2 Z^\mu Z_\mu \\ &+ g\left(\frac{v g}{2}\right) H W_+^\mu W_\mu^- + g \frac{v g}{4\cos^2(\theta_w)} H Z^\mu Z_\mu \\ &+ \frac{g^2}{4} H^2 W_+^\mu W_\mu^- + \frac{g^2}{4\cos^2(\theta_w)} H^2 Z^\mu Z_\mu + \text{const.} . \end{aligned} \tag{2.39}$$

The second line directly enables to read of the masses of the $W^\pm$- and Z^0-bosons[3],

$$m_{W^\pm} = \frac{1}{2} v g \qquad \text{and} \qquad m_{Z^0} = \frac{v g}{2\cos(\theta_w)} = \frac{m_{W^\pm}}{\cos(\theta_w)} . \tag{2.40}$$

Therefore, the ratio of the masses of the $W^\pm$- and Z^0-bosons depends only on the weak mixing angle.
Furthermore, the mass of the Higgs is given by

$$m_H = \sqrt{2\lambda} v . \tag{2.41}$$

Additionally, the Lagrangian density contains cubic (HVV) and quartic ($HHVV$) terms, which describe interactions between the Higgs boson and massive gauge bosons (V). The coupling strength of the cubic and quartic interactions is proportional to the mass of the gauge bosons, $m_{w^\pm}$ or m_{Z^0}, and the mass of the Higgs boson, m_H, respectively. There are also cubic (H^3) and quartic (H^4) Higgs-boson self-interaction terms in the Lagrangian density. All these couplings predicted by the spontaneous symmetry breaking enable to measure the relation between gauge boson masses and coupling strengths.

[3] A mass term has generally a form of $\frac{1}{2} m^2 X^\mu X_\mu$ for a neutral particle and $m^2 X^+ X^-$ for a charged particle, where X denotes the field associated with the particle.

Up to now only the gauge bosons and the Higgs boson acquire a mass. Fermion masses can be included in the theory by introducing a new coupling, which has to be invariant under the $SU(2)_{L,I_w} \times U(1)_Y$. The new coupling is called *Yukawa coupling* and describes the coupling between left-handed fermion $SU(2)_{L,I_w}$-doublets, right-handed fermion $U(1)_Y$-singles, and the Higgs-doublet.
For first-generation leptons the Lagrangian density of the Yukawa coupling is given by

$$\mathcal{L}^{\text{lep,1}}_{\text{Yukawa}} = -g_e(\overline{\nu}_e, \overline{e})_L \Phi e_R + h.c.\,, \tag{2.42}$$

where $h.c.$ denotes the corresponding hermitian conjugated term and g_e the coupling strength. Because of the form of the Higgs doublet Φ as defined in Eq. (2.35) only the electron acquires a mass, the neutrino remains massless.
Before the Yukawa coupling for quarks can defined, first an additional charge conjugated Higgs doublet Φ_C needs to be introduced, to enable couplings to quarks with $I_3 = \frac{1}{2}$,

$$\Phi_C(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} v + H(x) \\ 0 \end{pmatrix} . \tag{2.43}$$

Now, the Lagrangian density for the Yukawa coupling for quarks of the first generation can be written down,

$$\mathcal{L}^{\text{quark,1}}_{\text{Yukawa}} = -g_d(\overline{u}, \overline{d})_L \Phi d_R - g_u(\overline{u}, \overline{d})_L \Phi_C u_R + h.c. \tag{2.44}$$

For fermions of the second and third generation the same approach can be used.
It turns out that the coupling strength g_f for fermions is directly proportional to the correspinding mass m_f of the fermion,

$$m_f = v \frac{g_f}{\sqrt{2}}\,. \tag{2.45}$$

If the excitation of the ground state is considered, the Lagrangian density of the Yukawa coupling to fermions can be written as

$$\mathcal{L}_{\text{Yukawa}} = -m_f \overline{f} f \left(1 + \frac{H}{v}\right) . \tag{2.46}$$

This is the last missing piece of the full Lagrangian density of the Standard Model,

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{QCD}} + \mathcal{L}_{\text{EW}} + \mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{Yukawa}}\,. \tag{2.47}$$

The only unknown parameter is λ, since all other paraters are fixed by the masses of the gauge bosons and fermions, which can be experimentally measured. Thus, λ needs to be determined by measuring the mass of the Higgs boson.

2.2. The Higgs Boson

The Higgs boson was predicted by the Higgs mechanism in Standard Model since the 1960s. It has no electric or color charge and has a spin-parity configuration of

$J^{CP} = 0^+$. Its mass is not predicted and con only be determined by experiments. The coupling strength is proportional to the mass of fermions and proportional to the squared mass of gauge bosons.
In 2012 the Higgs boson was observed with the ATLAS and CMS experiments at CERN [1, 2] with a mass of $m_H = 125\,\text{GeV}$. Due to its short lifetime of about $10^{-22}\,\text{s}$ it cannot be observed directly, but only via its decay products. Because the cross-section of the Higgs boson is several magnitudes smaller than the ones from other processes produced at the LHC, as can be seen in Fig. 2.2, refined analysis strategies are needed. In this section the production of the Higgs boson at the LHC is discussed, followed by an overview of the decay modes of the Higgs boson. Measurements of the properties of the Higgs boson are discussed in Section 2.3.

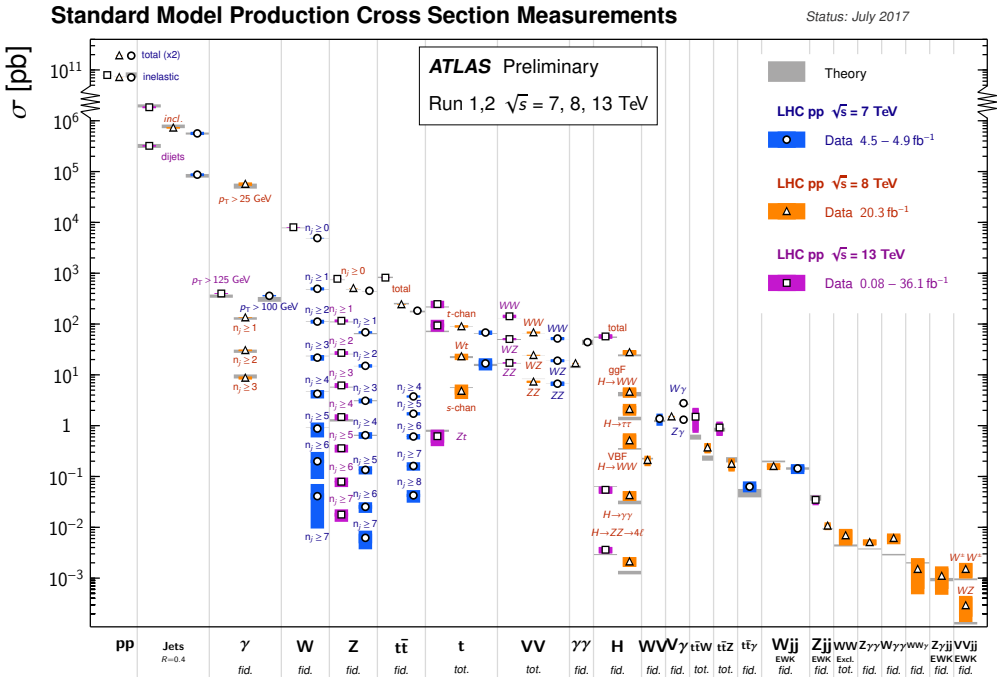


Figure 2.2.: Overview of inclusive and fiducial SM cross-sections measured during Run-1 and Run-2 and compared to theory predictions [44].

2.2.1. Higgs Boson Production in Proton–Proton Collisions

In proton–proton collisions at the LHC the constituents of the protons can be described as free, charged, point-like particles, which are called *partons*. The possibility to find a parton with a momentum fraction x of the total momentum of the proton is given by the parton distribution function (PDF), $f(x_i, Q^2)$, which depends on the squared momentum transfer, Q^2. To calculate the cross-section of the production of a particle X in proton–proton collisions the cross-section at parton level $\hat{\sigma}_{ij\to X}$ has to be weighted with the PDFs and all possible momentum

fractions need to be considered, as prescribed by the factorization theorem [45]. Mathematically speaking this is a convolution of the PDFs with the partonic cross-section,

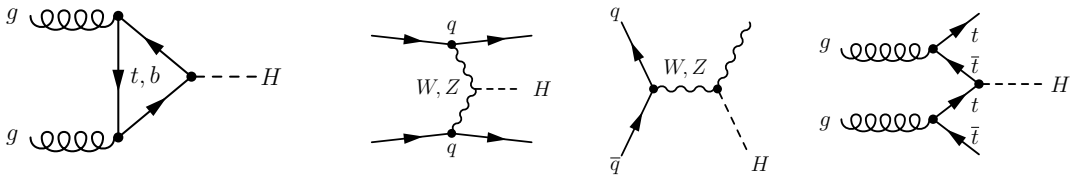
$$\sigma_X = \sum_{i,j} \int_0^1 \mathrm{d}x_i \int_0^1 \mathrm{d}x_j \, f_i\left(x_i, \mu_F^2\right) f_j\left(x_j, \mu_f^2\right) \hat{\sigma}_{ij\to X}(\alpha_s, \mu_R^2)\,, \tag{2.48}$$

where μ_F is the factorization scale and μ_F the renormalization scale. The general expression for the partonic cross-section is [18]

$$\hat{\sigma}_{ij\to X} = \frac{1}{F} \int \mathcal{M}\left(ij \to X\right) \mathrm{d}\Phi\,, \tag{2.49}$$

with the matrix element $\mathcal{M}$ which describes the transition probability of the initial state ij to the final state X, the particle flux F, and the phase-space factor $\mathrm{d}\Phi$ depending on the kinematics of the collision.

The Higgs boson can be produced in multiple ways, which vary in cross-section and phenomenology. In Fig. 2.3 the leading order (LO) Feynman diagrams are shown for the dominant production modes at the LHC.

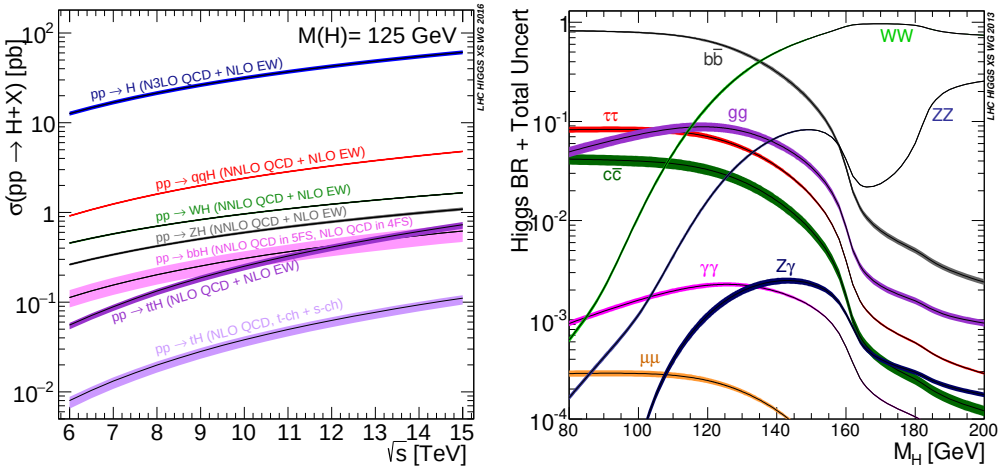


(a) gluon–gluon fusion **(b)** vector boson fusion **(c)** Higgs-Strahlung **(d)** associated production with top-quarks

Figure 2.3.: Feynman diagrams of the dominant production modes of the Higgs boson at the LHC. The cross-section decreases from left to right.

The production mode with the highest cross-section is gluon–gluon fusion (ggF). This is caused by the high contribution of the gluon PDF in protons for small momentum fractions x, which enables a quark loop producing a Higgs boson. Because coupling strength of the Higgs boson is proportional to the mass of the interaction particle, top and bottom quarks contributions dominate in the quark loop. At leading order only the Higgs boson is produced, therefore it has no transverse momentum. However, at higher orders final state QCD radiation is possible, which acts as a recoil partner for the Higgs boson. This is important for measurements to reduce background contributions.
The cross-section of the vector-boson fusion (VBF) production mode is one order below the one of gluon–gluon fusion. Here two initial state quarks radiate a Z^0 or $W^\pm$ boson. The bosons annihilate and produce a Higgs boson. The two final state quarks provide a characteristic signature, which is defined by a high mass of

the dijet system and a large separation of the two jets in the pseudorapidity η as defined in Eq. (3.4).
Another production mode is the so-called Higgs-Strahlung, where one weak boson created by the annihilation of a quark–antiquark pair radiates a Higgs boson.
The Higgs boson production associated with a top-quark pair has a suppressed cross-section compared with the other production modes. Because of the large mass of the top quark a high invariant mass is required, which reduces the available phase space.
A distribution of the cross-sections for different Higgs-boson production-modes as a function of the center-of-mass energy $\sqrt{s}$ is shown in Fig. 2.4a. The values of the cross-sections of the most dominant production modes of the Higgs boson at $\sqrt{s} = 13\,\text{GeV}$ and corresponding uncertainties are listed in Table 2.4.



(a) Production cross-sections. **(b)** Decay modes.

Figure 2.4.: Production cross-sections of the Higgs boson as a function of the center-of-mass energy $\sqrt{s}$ (a) [46] and branching ratios of the Higgs boson decay-modes depending in the mass m_H of the Higgs boson (b) [47].

2.2.2. Decay Modes of the Higgs Boson

The coupling strengths m_f/v and m_V^2/v of the Higgs-boson coupling to fermions f and gauge bosons V are proportional to the masses of the interacting particles. Thus, the branching ratio (BR), which is defined as the fraction of partial decay width to the total decay width,

$$BR(H \to X) = \frac{\Gamma(H \to X)}{\Gamma_{\text{total}}}\,, \tag{2.50}$$

increasing for higher masses of the decaying particles.
The dominant decay channels for fermions are $H \to b\bar{b}$, $H \to \tau^+\tau^-$, $H \to c\bar{c}$, and $H \to \mu^+\mu^-$. Furthermore, the Higgs boson can decay directly into massive gauge

Table 2.4.: Cross-sections for different production modes of the Higgs boson with different uncertainties for proton–proton collisions at $\sqrt{s} = 13\,\text{GeV}$ and a mass of the Higgs boson of $m_H = 125\,\text{GeV}$. Additionally the fraction with respect to the total cross section of the Higgs boson, $\sigma_i/\sigma_{\text{tot}}$, is given [46].

Mode	σ/pb	$\delta_{\text{QCD scale}}$	δ_{PDF}	δ_{α_s}	$\sigma_i/\sigma_{\text{tot}}$
ggF	48.58	$^{+4.6\,\%}_{-6.7\,\%}$	$\pm 1.9\,\%$	$\pm 2.6\,\%$	88 %
VBF	3.782	$^{+0.4\,\%}_{-0.3\,\%}$	$\pm 2.1\,\%$	$\pm 0.5\,\%$	6.7 %
WH	1.373	$^{+0.5\,\%}_{-0.7\,\%}$	$\pm 1.7\,\%$	$\pm 0.9\,\%$	2.5 %
ZH	0.8839	$^{+3.8\,\%}_{-3.1\,\%}$	$\pm 1.3\,\%$	$\pm 0.9\,\%$	1.60 %
ttH	0.5071	$^{+5.8\,\%}_{-9.2\,\%}$	$\pm 3.0\,\%$	$\pm 2.0\,\%$	0.92 %

bosons, $H \to WW^*$ and $H \to ZZ^*$. Using a heavy quark loop and W-boson loop, it can also decay into massless gauge bosons, $H \to gg$ and $H \to \gamma\gamma$.

The branching ratios depending on the mass of the Higgs boson are shown in Fig. 2.4b. The corresponding values for a Higgs boson with a mass $m_H = 125\,\text{GeV}$ are listed in Table 2.5. At the LHC the two most dominant decay modes are the decay into a bottom-quark pair with a branching ratio of 58.24 % and the decay into a pair of W-bosons with a BR of 21.37 %. The focus of this analysis is on the decay into a pair of τ-leptons, which has a branching ratio of 6.272 %.

Table 2.5.: Branching ratios for various decay modes of the Higgs boson with a mass of $m_H = 125\,\text{GeV}$ with corresponding theoretical uncertainties (THU) and parametric uncertainties from the quark masses (PU(m_q)) and the strong coupling constant (PU(α_S)) expressed in percent [46].

Decay channel	Branching Ratio [%]	THU [%]	PU(m_q) [%]	PU(α_s) [%]
$b\bar{b}$	58.24	$^{+0.65}_{-0.65}$	$^{+0.72}_{-0.74}$	$^{+0.78}_{-0.80}$
WW	21.37	$^{+0.99}_{-0.99}$	$^{+0.99}_{-0.98}$	$^{+0.66}_{-0.63}$
gg	8.187	$^{+3.40}_{-3.41}$	$^{+1.12}_{-1.13}$	$^{+3.69}_{-3.61}$
$\tau^+\tau^-$	6.272	$^{+1.17}_{-1.16}$	$^{+0.98}_{-0.99}$	$^{+0.62}_{-0.62}$
$c\bar{c}$	2.891	$^{+1.20}_{-1.20}$	$^{+5.26}_{-0.98}$	$^{+1.25}_{-1.25}$
ZZ	2.619	$^{+0.99}_{-0,99}$	$^{+0.99}_{-0.98}$	$^{+0.66}_{-0.63}$
$\gamma\gamma$	0.2270	$^{+1.73}_{-1.72}$	$^{+0.93}_{-0.99}$	$^{+0.61}_{-0.62}$
$Z\gamma$	0.1533	$^{+5.71}_{-5.71}$	$^{+0.98}_{-1.01}$	$^{+0.58}_{-0.65}$
$\mu^+\mu^-$	0.021 76	$^{+1.23}_{-1.23}$	$^{+0.97}_{-0.99}$	$^{+0.59}_{-0.64}$

2.3. Measurements of the Higgs Boson at the LHC

This section gives an overview of recent measurements of Higgs-boson properties during Run-1 and Run-2 of the LHC.

2.3.1. Discovery

The observation of a new particle with a mass of 125 GeV in the search for a SM Higgs-boson was announced by both the ATLAS and CMS experiment on 4$^{\text{th}}$ of July 2012 [1,2]. The observed significance was 5.9σ and 5.0σ for ATLAS and CMS, respectively, which is passing the threshold of 5σ needed for an observation. The p_0 values as a function of the mass of the Higgs boson can be seen in Fig. 2.5. Results from the $H \to \gamma\gamma$, $H \to ZZ$, $H \to WW$, $H \to \tau\tau$, $H \to b\bar{b}$ decay channels were used for the observation.[4]

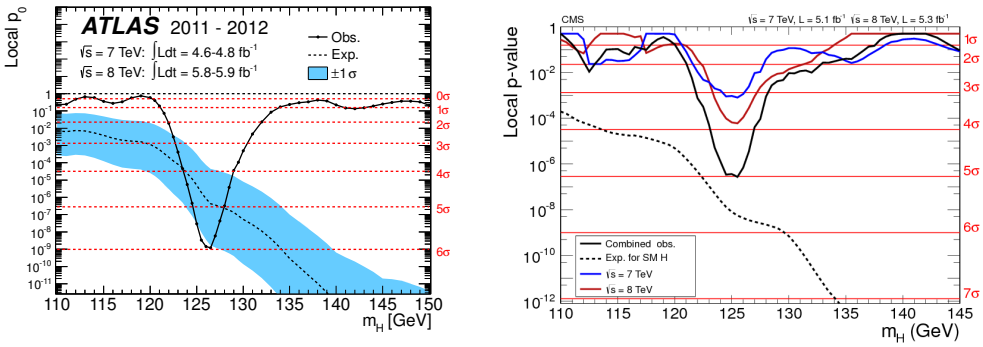


Figure 2.5.: Values for p_0 depending on the mass of the Higgs boson, m_H, from the observation of the Higgs boson during Run-1 with the ATLAS (left, [1]) and CMS (right, [2]) experiments. The black dashed lines show the expected values. The corresponding significances are indicated by red lines.

2.3.2. Measurements during Run-1

During Run-1 several properties of Higgs boson were measured, like the mass, signal strength, decay width, as well as spin and CP properties.

Mass measurement

The mass of the Higgs boson was determined individually by the ATLAS and CMS experiment and in a combined measurement using the full Run-1 dataset. For the mass measurement only the $H \to \gamma\gamma$ and $H \to ZZ \to 4\ell$ channels were used, since

[4] The $H \to b\bar{b}$ channel was only used by CMS.

other channels provide a worse mass resolution or do not have direct access to the mass of the Higgs boson. The combined result is [3]

$$m_H = (125.09 \pm 0.21(\text{stat.}) \pm 0.11(\text{sys.}))\,\text{GeV}\,, \tag{2.51}$$

individual results are shown in Fig. 2.6.

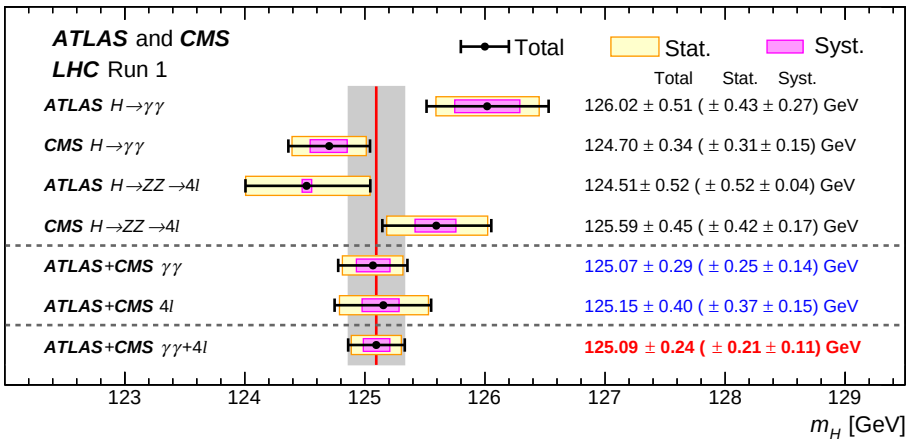


Figure 2.6.: Measurement of the mass of the Higgs boson in the $H \to \gamma\gamma$ and $H \to ZZ \to 4\ell$ decay channels during Run-1 with the ATLAS and CMS detector [3].

Signal strength

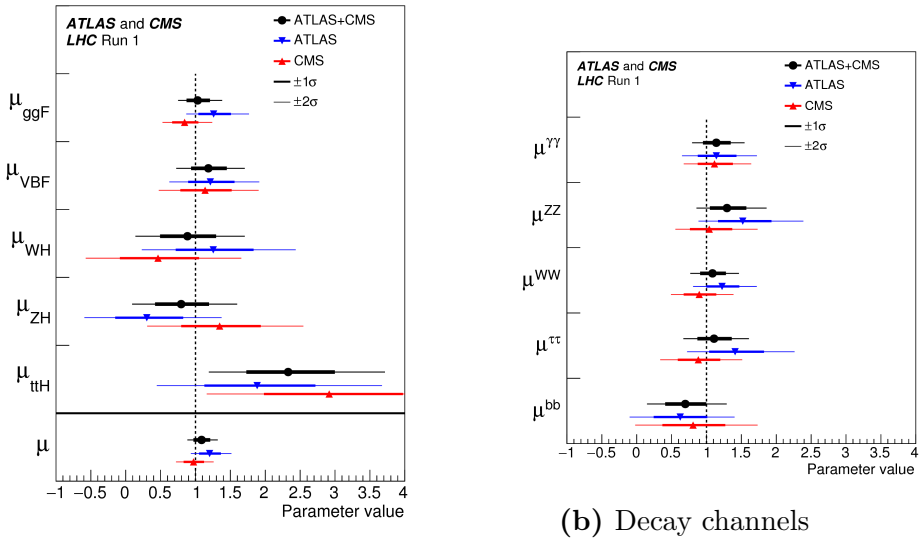
The measured cross section of a process is usually not given directly, but the ratio of the measured cross section to the prediction of the SM is used,

$$\mu = \frac{\sigma_{\text{experimental}}}{\sigma_{\text{SM}}}\,. \tag{2.52}$$

This ratio is the so-called *signal strength* and is denoted with μ. This makes it very easy to check if the measured cross section is in agreement with the standard model value for different processes, which would be always indicated by $\mu = 1$.
The signal strength of the Higgs boson can be determined for different production modes and decay channels, as shown in Fig. 2.7. Since the signal strength is the product of the production and decay signal strength, the decay cross sections are assumed to be ± 1 for the measurement of the production cross sections and vice versa. All signal strengths agree within uncertainties with the predicted SM value of $\mu = 1$. The combined signal strength is [48]

$$\mu = 1.09 \pm 0.07(\text{stat.})^{+0.09}_{-0.08}(\text{sys.})\,, \tag{2.53}$$

which matches with the SM expectation within the uncertainties.



(a) Production modes.

(b) Decay channels

Figure 2.7.: Values of the signal strength of the Higgs boson in different production modes (a) and decay channels (b) obtained in a combined measurement of the ATLAS and CMS collaborations [48].

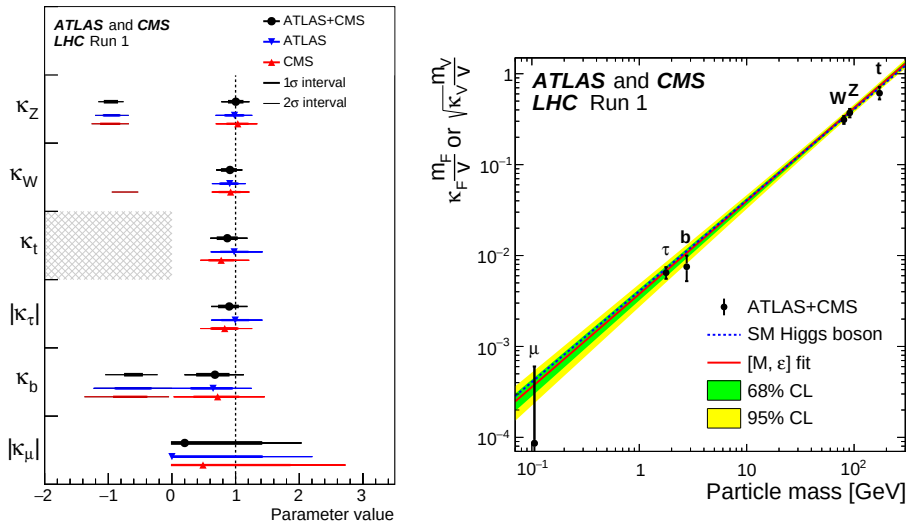
Additionally, the LO-coupling modifiers

$$\kappa_i = \frac{g_i}{g_{,\mathrm{SM}}} \tag{2.54}$$

can be measured, where g_i is the measured coupling strength and $g_{i,\mathrm{SM}}$ the prediction of the Standard Model. Here, the assumption that beyond the Standard-Model contributions are not present in loops and decay needs to be made. The normalized coupling strengths are shown in Fig. 2.8a. All values are in agreement with the SM expectation of $\kappa_i = 1$.
Furthermore, the reduced coupling strengths $\kappa_F \frac{m_F}{v}$ for fermions and $\sqrt{\kappa_V}\frac{m_V}{v}$ for gauge bosons can be calculated. The vacuum expectation value is denoted by v. If all the reduced coupling strengths are depicted as a function of the particle mass, a linear dependence is predicted by the Standard Model. All values agree within uncertainties with the SM expectation, as can be seen in Fig. 2.8b

Decay width

The Standard Model prediction for the decay width of a Higgs boson with a mass of 125 GeV is $\Gamma_\mathrm{H} = 4\,\mathrm{MeV}$ [49]. This makes the direct measurement of the decay width nearly impossible, since the detector resolution for the $H \to \gamma\gamma$ and $H \to ZZ \to 4\ell$ channels is around three magnitudes large.
A direct measurement of the decay width of the Higgs boson yielded an upper limit of around 1 GeV [50, 51].
Refined analysis techniques are needed for a better limit. One option is to measure the flight distance of the Higgs boson in the detector, from which the lifetime



(a) LO-coupling modifiers

(b) Reduced coupling strength

Figure 2.8.: LO-coupling modifiers of the Higgs boson to different particles (a) and reduced coupling strengths of the Higgs boson to different particles as the function of the particle mass (b) determined in a combined measurement of the ATLAS and CMS collaboration. The prediction of the Standard Model is indicated by the dashed line [48].

and subsequently the decay width can be calculated. This was done by the CMS collaboration and resulted in a lower limit of $\Gamma_H > 3.5 \times 10^{-9}\,\text{MeV}$ [52].
Another possibility is to exploit the production of off-shell production $gg \to H \to VV$ channel. The ratio between the on- and off-shell signal strengths can be used to determine the decay width of the Higgs boson [53],

$$\frac{\mu_{\text{off-shell}}}{\mu_{\text{on-shell}}} = \frac{\Gamma_H}{\Gamma_H^{\text{Theory}}}\,. \tag{2.55}$$

Both the WW and ZZ decay channel can be used. Here, the assumption is made that the off-shell cross-section does not depend on the partonic center-of-mass energy and that the background prediction is not modified by new physics. The combined limit of those two channels from the CMS collaboration is $\Gamma_H < 13\,\text{MeV}$ [53, 54] and from the ATLAS collaboration $\Gamma_H < 22.7\,\text{MeV}$ [55] at 95 % confidence level (CL).

Spin and CP properties

In the Standard Model the spin and CP properties of the Higgs boson are predicted to be $J^{CP} = 0^+$, i.e. a Higgs boson with spin zero and CP even nature is expected. Both ATLAS and CMS could confirm this prediction and exclude models with other values of J^{CP} at more than 99.9 % CL in measurements using the $H \to \gamma\gamma$ and $H \to ZZ \to 4\ell$ and $H \to WW \to 2\ell 2\nu$ decay channels [56, 57]. Especially the

CMS collaboration tested a multitude of alternative spin-2 models, as shown in Fig. 2.9.

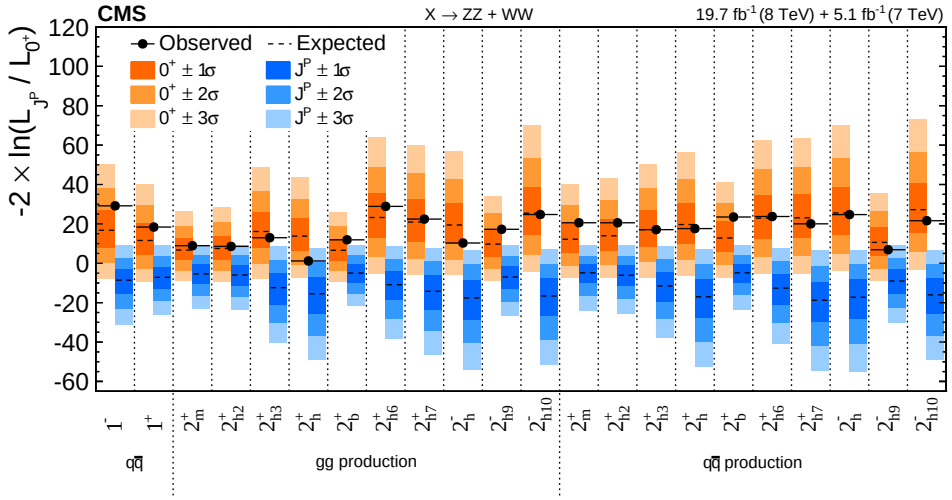


Figure 2.9.: Observed and expected values of the test statistic discriminating between the SM prediction and other hypothesis from the spin and parity analysis of the CMS collaboration [57].

Those measurements do not give information about the sign of the CP operator. The state $J^{CP} = 1^+$ only indicates that the Higgs boson is invariant under CP transformations. However, this does not fix the value of the eigenstate of the CP operator. The possibilities are either $+1$ or -1, which are also called CP even or CP odd, respectively, or a mixing between those states.
The mixing between CP-even and CP-odd eigenstates can be measured using the so-called *Optimal Observable*, which was already used during Run-1 in the $H \to \tau\tau$ decay channel analysis [58]. Here, the CP-mixing parameter $\tilde{d}$ provided by the Optimal Observable is directly related to the CP-even nature of the Higgs boson in the VBF production mode. A CP-even Higgs-boson results in $\tilde{d} = 0$ while deviations from zero of the CP-mixing parameter would indicate a CP-violating nature of the Higgs boson.
The results of the measurements in [58] are in agreement with the SM prediction of $\tilde{d} = 0$ and exclude values of $\tilde{d}$ outside of $[-0.11, 0.05]$ at a confidence level of 68 %, as can be seen in Fig. 2.10.
Additionally, the CP mixing nature of the Higgs boson can also be measured in the $H \to WW$ and $H \to ZZ$ decay modes [59, 60].

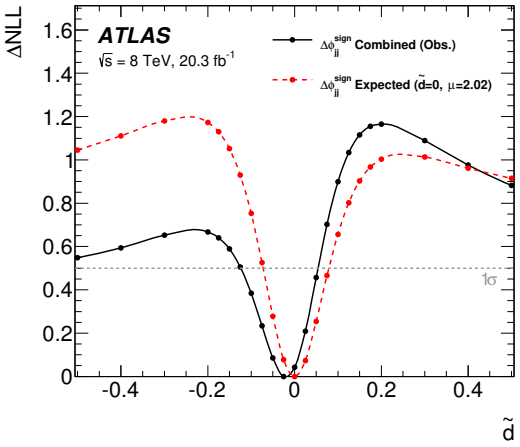


Figure 2.10.: The negative log likelihood (ΔNNL) as a function of the CP-mixing parameter $\tilde{d}$ obtained by the Optimal Observable method in the VBF $H \to \tau\tau$ decay channel during Run-1 with the ATLAS detector. The minimum of the ΔNNL curve corresponds to the best fit value. The 68 % confidence level is indicated by the gray dashed line [58].

2.3.3. Measurements during Run-2

In 2015 the second data-taking period, Run-2, started with an increased center-of-mass energy of 13 TeV. In the following a few results from the ATLAS experiment using the 2015 and 2016 dataset are presented.

Total cross section

Due to the changing center-of-mass energy the dependence of the total cross section on $\sqrt{s}$ can be investigated. For this the $H \to \gamma\gamma$ and $H \to ZZ \to 4\ell$ decay channels are used. The cross section is measured for a center-of-mass energy of 7 (2011), 8 (2012), and 13 TeV (2015+2016), with a dataset corresponding to 4.5, 20.3, and 36.1 fb^{-1}, respectively [61]. The results are shown in Fig. 2.11. Due to the increased cross-section from Run-1 to Run-2, the total cross section is increased by a factor of approximately 2, agreeing with the SM prediction.

Signal strength

The signal strength was measured again with the 2015 and 2016 dataset from Run-2 corresponding to 36.1 fb^{-1} with the ATLAS detector. In the measurement the $H \to \gamma\gamma$ and $H \to ZZ \to 4\ell$ decay channels were used, yielding a combined signal strength of [61]

$$\mu = 1.09 \pm 0.09(\text{stat.})^{+0.06}_{-0.05}(\text{exp.})^{+0.06}_{-0.05}(\text{theo.})\,. \tag{2.56}$$

This is in agreement with the prediction of the Standard Model and previous measurements.

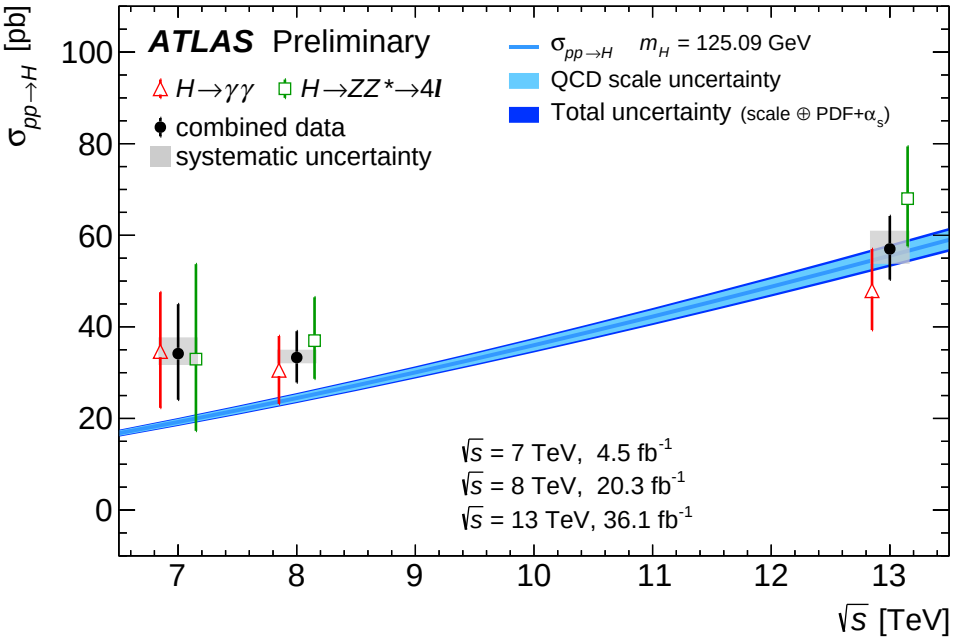


Figure 2.11.: Measurement of the total cross section as a function of the center-of-mass energy in the $H \to \gamma\gamma$ and $H \to ZZ \to 4\ell$ decay channels with the ATLAS detector [61].

Mass measurement

The mass of the Higgs boson was also determined again using the $H \to \gamma\gamma$ and $H \to ZZ \to 4\ell$ decay channels. In the measurement the full 2015 and 2016 dataset corresponding to 36.1 fb^{-1} is used. The result is [62]

$$m_H = (124.98 \pm 0.19(\text{stat.}) + 0.21(\text{sys.}))\,,\text{GeV}\,, \tag{2.57}$$

which is in agreement of the combined measurement of Run-1.

3. Experimental Setup

3.1. The Large Hadron Collider

The Large Hadron Collider (LHC) [63] is currently the worlds largest and most powerful proton and heavy ion accelerator. It is located at CERN (Conseil Européen pour la Recherche Nucléaire) near Geneva.
The LHC was constructed between 1998 and 2008 inside the circular, 27 km long tunnel of the former Large Electron–Positron (LEP) Collider, which was shutdown in 2000. The tunnel is located between 50 and 175 meters below ground level and crossing the France–Switzerland border. Both protons and heavy ions can be accelerated in two beam pipes in opposite direction. In proton–proton collisions both beams can contain up to 2808 bunches which contain 10^{11} protons each. The time distance between the bunches is 25 ns. To bend the proton beams 1232 superconducting dipole magnets are used, which can generate a magnetic field of up to 8.3 T. Additional 392 quadrupole magnets are used to focus the beams.
The beams are collided at four *interaction points* (IP), where the four major experiments ATLAS [64], CMS [65], LHCb [66], and ALICE [67] are located. ATLAS and CMS are multipurpose detectors and are used to perform a wide range of measurements and searches. The focus of LHCb are interactions of B-hadrons. ALICE is specialized for measurements of heavy-ion collisions. Fig. 3.1 shows a schematic overview of the LHC and its experiments. The ATLAS experiment is discussed in detail in Section 3.2.
The number of events per second which are generated in LHC collisions for a particular process is given by

$$N_{\text{proc}} = \mathcal{L}\sigma_{\text{proc}} \,, \tag{3.1}$$

where σ_{proc} is the cross-section for thiss process and $\mathcal{L}$ the instantaneous luminosity. The instantaneous luminosity is a quantity of the LHC and depends only the parameters of the beams. For bunches with a Gaussian shape distribution it can be written as [63]

$$\mathcal{L} = \frac{N_b^2 n_b f_{\text{ref}} \gamma_r}{4\pi\epsilon_n \beta^*} F \,, \tag{3.2}$$

where N_b is the number of particles in each bunch, n_b the number of bunches per beam, f_{ref} the revolution frequency of the particles, γ_r the relativistic gamma factor, ϵ_n the normalized transverse beam emittance, β^* the beta function at the collision point, and F the geometric luminosity reduction factor due to the crossing angle of the beams at the interaction point. The LHC is designed to collide protons with an instantaneous luminosity of up to $\mathcal{L} = 10^{34}\,\text{cm}^2\,\text{s}^{-1}$ and a beam energy of up

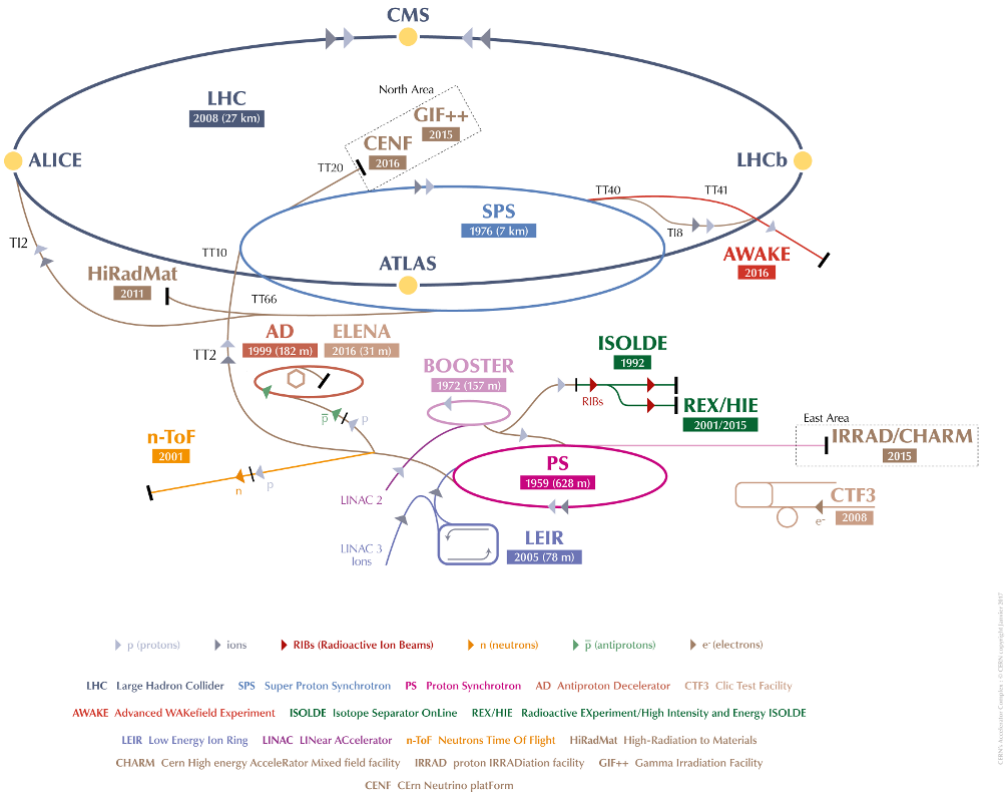


Figure 3.1.: The CERN accelerator complex, including the LHC and its preaccelerators. The four main experiments (ATLAS, CMS, LHCb, ALICE) are shown as a yellow dot [68].

to 7 TeV, which results in a collision with a center-of-mass energy of $\sqrt{s} = 14\,\text{TeV}$. Due to the high center-of-mass energies several preaccelerators which are shown in Fig. 3.1 are needed to accelerate the particles to the desired velocity.
The first data-taking period, labeled Run-1, was in 2011 and 2012 with center-of-mass energies of 7 and 8 TeV, respectively. A total amount of $28.3\,\text{fb}^{-1}$ of data was provided by the LHC [69]. The second data-taking period, Run-2, started in 2015 and will continue until 2018 with a center-of-mass energy of 13 TeV. For the years 2015 and 2016 data corresponding to an integrated luminosity of $\mathcal{L}_{\text{int}} = 42.7\,\text{fb}^{-1}$ were produced [70].

3.2. The ATLAS Experiment

The ATLAS (A Toroidal LHC ApparatuS) [64] detector is a general-purpose detector located at LHC *Point 1* about 100 meters below ground level. It is designed to measure properties of SM particles and processes with a high precision and for the discovery of new particles in hadron collisions at high energies.
The detector is of cylindrical shape with a length of 40 meters and a diameter of 25 meters, and weighs around 7000 metric tons. It is forward-backward symmetric with respect to the interaction point.

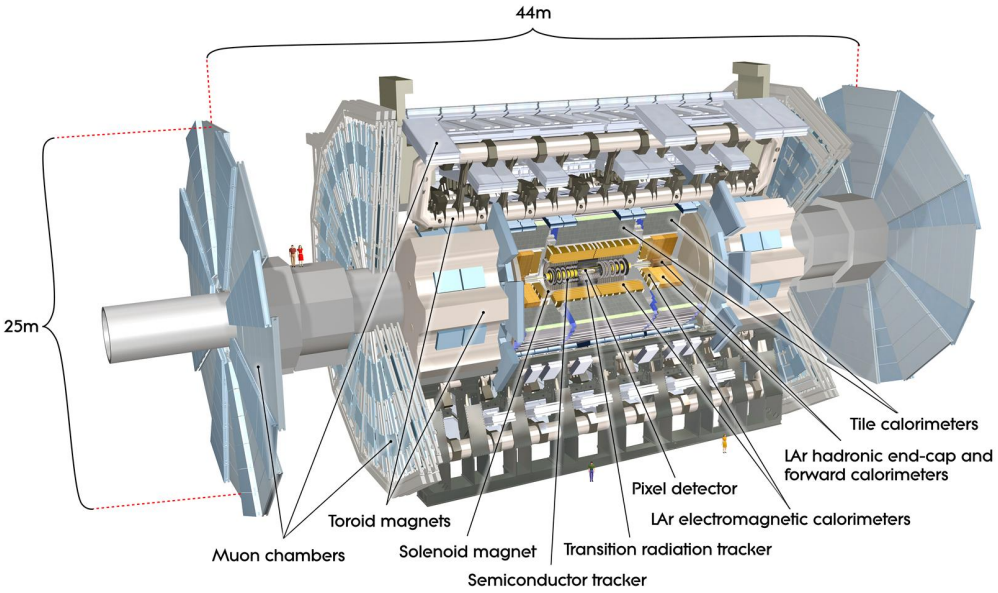


Figure 3.2.: Overview of the ATLAS detector and its subsystems [64].

The ATLAS detector consists of several subdetector systems, which are illustrated in Fig. 3.2. The innermost system is the inner detector (ID), which is used to measure trajectories of charged particles, whose flight path is bent by a 2 T magnetic field, generated by a superconducting solenoid. Next are the electromagnetic (ECal) and hadronic (HCal) calorimeters, which measure energy depositions with liquid-argon and scintillator-tile technology. The outermost part is the muon spectrometer (MS) which measures trajectories of muons. An additional magnet system composed of large toroid magnets which gives the ATLAS detector its distinct look is used to bend the trajectories of the particles again. The different subdetectors are discussed in more detail in Sections 3.2.2 to 3.2.4. The general resolution goals and pseudorapidity coverage of the individual subdetectors is listed in Table 3.1.
Due to the high luminosity provided by the LHC a lot of additional inelastic scattering events occur during a bunch crossing. Fast readout electronics are needed to select events which are relevant to analyze. A sophisticated trigger system as discussed in Section 3.2.5 is used to reduce the amount of events.

3.2.1. Nomenclature

The ATLAS experiment uses a right-handed coordinate system where the origin is located at the nominal interaction point. The x-axis points towards the center of the LHC ring and the y-axis points upwards. Therefore, the z-axis points counterclockwise, if viewed from above, in beam direction. The polar angle θ is defined with respect to the z-axis and the azimuthal angle ϕ is measured in the x–y plane.
Usual variables at hadron colliders are the energy and momentum of a particle

Table 3.1.: Resolution goals and pseudorapidity coverage of the subsystems of the ATLAS detector. Numbers for energy and transverse momentum are in GeV. The notation $a \oplus b = \sqrt{a^2 + b^2}$ is used. [64]

Subdetector	Required Resolution	η-coverage	
		Measurement	Trigger
Inner Detector	$\sigma_{p_\mathrm{T}}/p_\mathrm{T} = 0.05\,\% p_\mathrm{T} \oplus 1\,\%$	$\lvert\eta\rvert < 2.5$	
Electromagnetic Calorimeter	$\sigma_E/E = 10\,\%/\sqrt{E} \oplus 0.7\,\%$	$\lvert\eta\rvert < 3.2$	$\lvert\eta\rvert < 2.5$
Hadronic Calorimeter			
barrel and end-cap	$\sigma_E/E = 50\,\%/\sqrt{E} \oplus 3\,\%$	$\lvert\eta\rvert < 3.2$	$\lvert\eta\rvert < 3.2$
forward	$\sigma_E/E = 100\,\%/\sqrt{E} \oplus 10\,\%$	$3.1 < \lvert\eta\rvert < 4.9$	$3.1 < \lvert\eta\rvert < 4.9$
Muon Spectrometer	$\sigma_{p_\mathrm{T}}/p_\mathrm{T} = 10\,\%$ at $p_\mathrm{T} = 1\,\mathrm{TeV}$	$\lvert\eta\rvert < 2.7$	$\lvert\eta\rvert < 2.4$

in the transverse plane, since they are independent on the boost of the system of the colliding particles in beam direction. The symbols E_T and p_T are used for the scalar quantities, respectively. A bold symbol like $\mathbf{p}_\mathrm{T}$ is used for vectorial quantities. The rapidity of an object is defined as

$$y = \frac{1}{2} \log \left(\frac{E + p_z}{E - p_z} \right) , \tag{3.3}$$

where p_z is the z-component of the momentum of the object. In the case of a relativistic or massless particle ($E \gg m$) the rapidity can be replaced with the pseudorapidity η, which is defined as

$$\eta = -\log \tan \frac{\theta}{2} , \tag{3.4}$$

which only depends on the polar angle θ. Differences of rapidity, Δy and $\Delta\eta$, are *Lorentz invariant* under boosts along the z-axis, which is convenient when working with objects originating from hadron collisions. This holds also true for the ΔR separation, a quantity which describes the angular separation of two objects in the η–Φ plane,

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\Phi)^2} . \tag{3.5}$$

3.2.2. Inner Detector

The ATLAS inner detector is used to measure the trajectories (tracks) and momentum of charged particles with a transverse momentum above $p_\mathrm{T} > 0.5\,\mathrm{GeV}$. Those tracks can be used to reconstruct the primary and secondary vertices. The ID has a cylindrical shape with a length of 6.2 m and diameter of 2.1 m. A 2 T strong magnetic field produced by the central solenoid magnet, which cover the ID, is used to bend the flight path of the particles. The inner detector consists of

several subsystems: the pixel detector, semiconductor tracker (SCT), and transition radiation tracker (TRT), as shown in Fig. 3.3. Due to the close proximity of the inner detector to the beam pipe and the interaction point, the detector material is exposed to huge amounts of radiation and high temperatures. Therefore, extra radiation-hard material is used for the detectors. Additionally, the pixel detector and SCT are cooled down to around $-7\,°\mathrm{C}$ to mitigate damages.

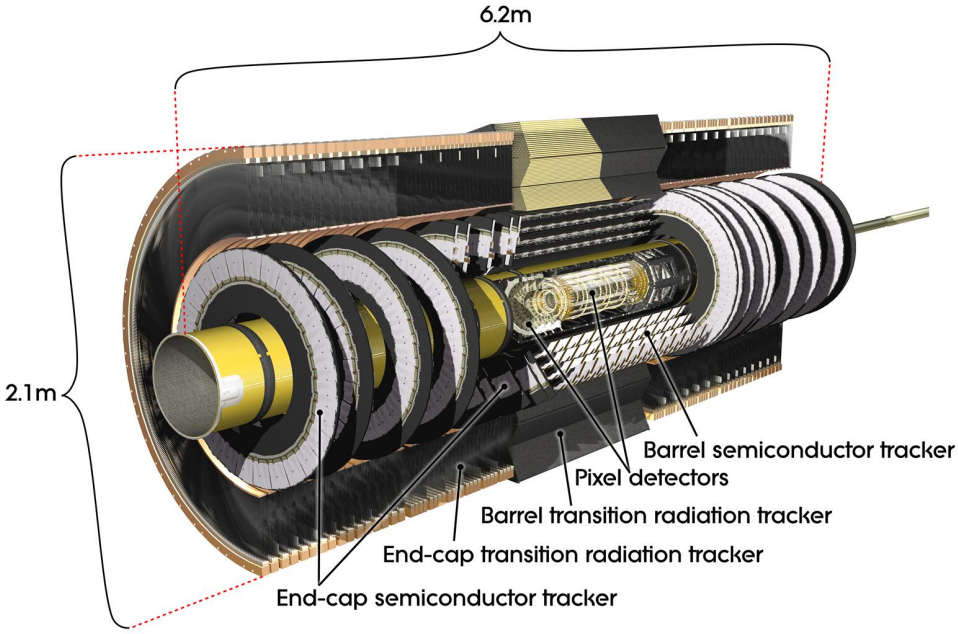


Figure 3.3.: Schematic overview of the inner detector with its submodules, the pixel detector, SCT, and TRT [64].

The pixel detector is closest to the beam pipe. It is composed of four barrel layers and two end-caps with each three discs. Both the barrel and end-cap layers are made of small silicon semiconductors called pixels. The layers are segmented in R–ϕ and z. The innermost barrel layer is the insertable B-layer (IBL), which was only added during the shutdown period between Run-1 and Run-2 [71]. The pixel detector covers the region of $|\eta| < 2.5$ and reaches a hit resolution of 10 µm in R–ϕ and 115 µm in z-direction.
The semiconductor tracker uses similar concepts as the pixel detector, silicon semiconductors are used as well. However, the used semiconductors are larger and have a strip-like geometry, which results in a worse resolution but a larger area which is covered compared to the pixel detector. The SCT is build out of four double layers of silicon strip detectors in the barrel part and nine layers in each of the end-caps. This ensures that every charged particles traverses at least four layers of detectors. For each double layer in the barrel region one set of the silicon strip modules is aligned to the beam axis and the other set is rotated by 40 mrad, which enables to measure the position along the beam axis. A hit resolution of 17 µm in the R–ϕ plane and 580 µm along the z-axis is achieved which a coverage

of $|\eta| < 2.5$.
The outermost part of the inner detector is the transition radiation tracker. It is made of gas-filled tubes, which are stabilized by carbon fibers. In the barrel region the tubes are aligned to the z-axis, for the end-caps they are positioned radially. Thus, only a position measurement in R–ϕ in the barrel region is possible, with a nominal hit resolution of 130 µm. The coverage is only $|\eta| < 2.5$. However, the TRT contributes substantially to the measurement of tracks, because of the high number of measured points per track (usually 36 points). Additionally, the TRT can be used for particle identification, since the transition radiation has an inverse dependence of the mass of the charged particle. Thus, the transition radiation is largest for electrons, which allows a discrimination from other particles.

3.2.3. Calorimeters

The ATLAS calorimeter system is placed around the solenoid which produces the magnetic field inside the ID. There are two types of calorimeters, the electromagnetic calorimeters and the hadronic calorimeter, as shown in Fig. 3.4. They cover the full ϕ range and a pseudorapidity of up to $|\eta| < 4.9$. Electrons and photons are absorbed in the electromagnetic calorimeter, which enables a measurement of their energy. The hadronic calorimter is used to determine the energy of hadrons. The missing transverse energy as defined in Eq. (5.1) is reconstructed by combining information from the two calorimeters, the tracking detectors, and the muon spectrometer.

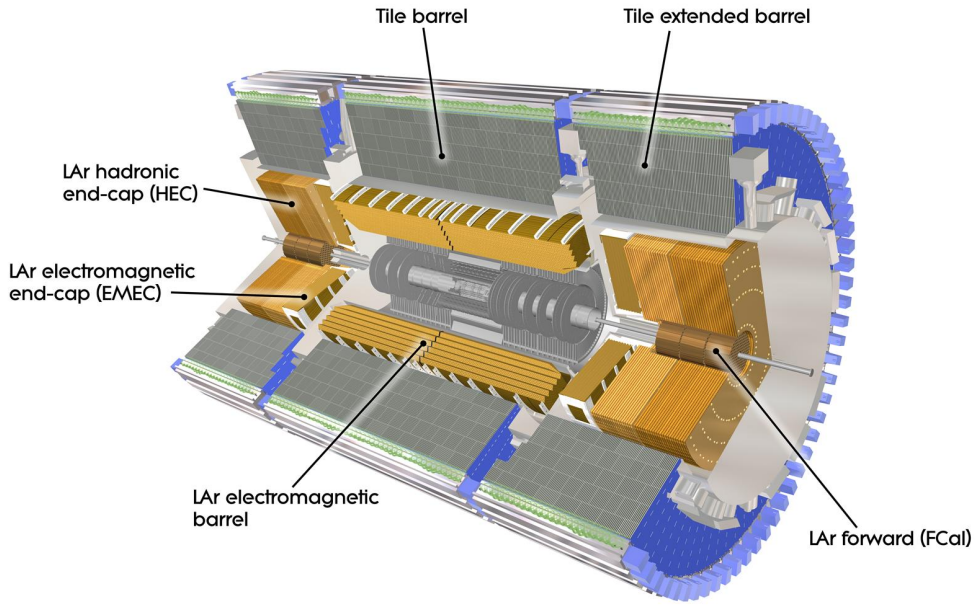


Figure 3.4.: Schematic overview of the electromagnetic and hadronic calorimeter [64].

Both calorimeters are *sampling calorimeters*, which means that they are made of alternating layers of active and absorbing material. It is essential for the calorimeters that a *punch-through* into the muon system is prevented. The calorimeters has to be thick enough that all energy depositions are contained within the calorimeters. This is important for a correct energy and $E_{\mathrm{T}}^{\mathrm{miss}}$ measurement.

Electromagnetic Calorimeter

The electromagnetic calorimeter is divided into the barrel region which covers $|\eta| < 1.475$ and the end-cap region within $1.375 < |\eta| < 3.2$, as can be seen in Fig. 3.4. It uses liquid argon (LAr) as the active sampling material and lead as absorber material.
The barrel part is composed of three layers of modules. The first layer has a fine segmentation in η which allows a precision measurement of the position of electrons and photons. The second and third layer have a coarser structure and are used to collect the bulk and tail of the electromagnetic showers, respectively. In all layers the calorimeter modules are arranged into an accordion-shaped structure to avoid gaps and enable coverage over the full ϕ range. The energy resolution in the barrel is shown in Fig. 3.5.
The two end-caps use the same accordion geometry. They are composed of an outer and inner wheel which cover $1.375 < |\eta| < 2.5$ and $2.5 < |\eta| < 3.2$, respectively. The inner wheel is made of three layers of modules, for the outer wheel only two layers are used.
To correct for energy which is lost due to the ID and solenoid a thin LAr sampling layer called the presampler is installed in front of the first layer in the region $|\eta| < 1.8$. For charged particles the track information of the ID can be matched to calorimeter cells within $|\eta| < 2.5$ to improve the measurement precision.

Hadronic Calorimeter

The hadronic calorimeter consists of three parts as shown in Fig. 3.4. It has a worse granularity and energy resolution than the electromagnetic calorimeter.
The tile calorimeter envelopes the electromagnetic barrel with a coverage of $|\eta| < 1.0$. It is supplemented by the two extended barrels which cover a range of $0.8 < |\eta| < 1.7$. Scintillator tiles are used as active material and steel plates as absorber.
The end-cap calorimeter consists of two wheels on each side which are used to cover the region of $1.5 < |\eta| < 3.2$. Here LAr is used as an active material and copper as an absorber. A distribution of the energy resolution in the hadronic end-cap calorimeter is shown in Fig. 3.5.
Compared to the other two parts of the hadronic calorimeter the forward detector can also be used to reconstruct photons and electrons, because different absorber materials are used. In the fist module copper is used as the absorber material, whereas the second and third layer use tungsten.
The first module is used to measure electromagnetic showers with copper as the absorber. The second and third module instead use tungsten as the absorber

material which allows the measurement of hadronic showers. All modules uses LAr as the active material. The forward calorimeter covers a range of $3.1 < |\eta| < 4.9$.

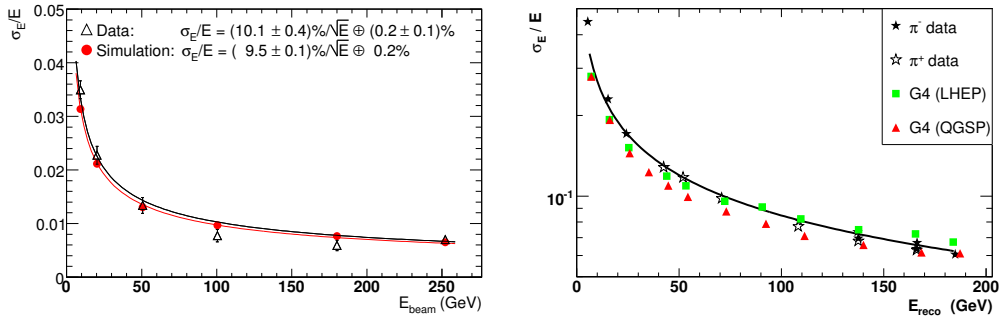


Figure 3.5.: Comparison between test-beam measurements and simulation of the energy resolution in the electromagnetic barrel calorimeter (left) and for pions in the hadronic end-cap calorimeter (right) [64].

3.2.4. Muon Spectrometer

The muon spectrometer is used to detect charged particles which pass the electromagnetic and hadronic calorimeters. These particles are mainly muons, since all other particles should deposit all their energy in the calorimeters. The MS is the outermost part of the ATLAS detector. A schematic overview can be found in Fig. 3.6. It is composed of three regions: the barrel region covering $|\eta| < 1.4$, the end-cap region which covers $1.6 < |\eta| < 2.7$, and the transition region, which covers the region $1.4 < |\eta| < 1.6$ between the two aforementioned regions. In the barrel region three layers of muon chambers are used, whereas four wheels perpendicular to the beam axis are installed in the end-cap region.

Large toroidal magnets produce the magnetic field needed for the momentum measurements. There are two different magnet systems, one for the barrel part and one for the end-caps, with strengths up to 2.5 T and 3.5 T, respectively. In the transition region the magnetic fields of both systems are used. In total 24 coils are used.

In most parts of the MS the trajectories of the muons are measured by monitored drift tubes, which provide a resolution of 35 µm per chamber. The one exception is the range of $2.0 < |\eta| < 2.7$ in the forward region, where cathode strip-chambers are used in the innermost layer. They provide a higher rate capability and time resolution. However, a spatial resolution of only 40 µm in the bending plane and 5 mm in the transverse plane is achieved.

The muon system also provides a trigger for particles in the range $|\eta| < 2.4$. Resistive plate chambers are used in the barrel region and thin gap chambers in the end-cap region, which achieve a response time of a few nanoseconds.

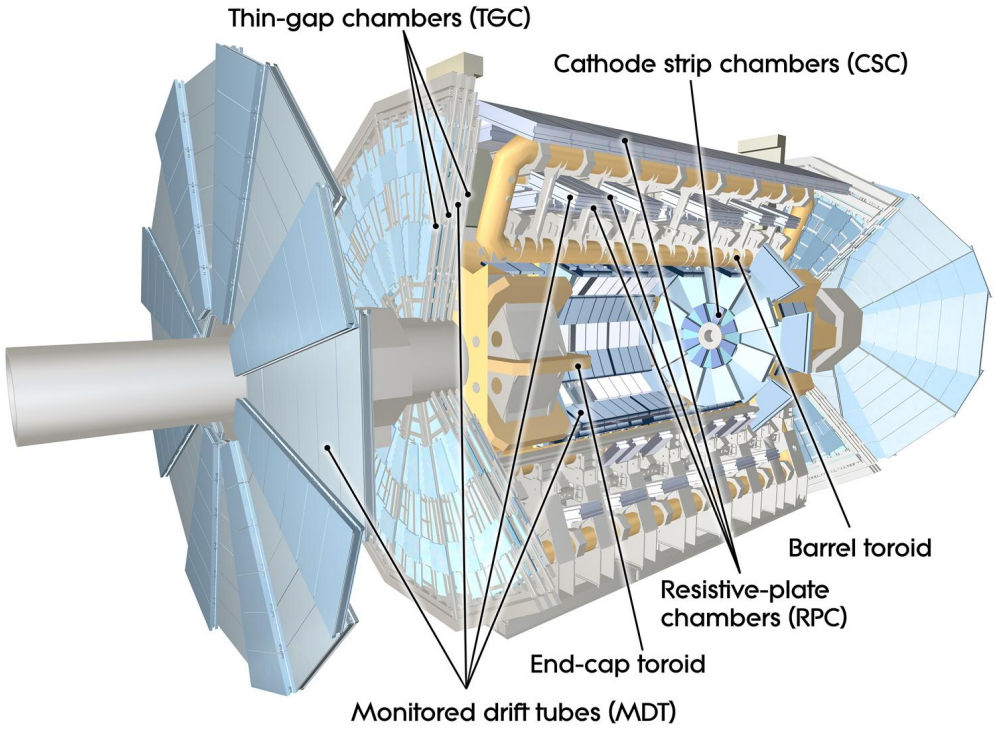


Figure 3.6.: Schematic overview of the ATLAS muon spectrometer with its different regions and the toroid system [64].

3.2.5. Trigger System

Particle bunches at the LHC have a time separation of 25 ns, therefore the rate of collisions is 40 MHz. However, only a small amount of the collision events can be recorded and further analyzed, due to the huge amount of data which is produced. A two level trigger-system is used to select events which are relevant for physics analyses. It is composed of the hardware-based first level trigger (L1) and the software-based high level trigger (HLT). An schematic overview is given in Fig. 3.7. The hardware-based L1 trigger uses coarse granularity information of calorimeters and the muon chambers provided by custom hardware to detect events where particles like electrons, τ-leptons, and jets have a high transverse energy. Events with large missing transverse energy or where muons have a large transverse momentum are also triggered. The decision time to accept an event is 2.5 µs and results in an event rate of 100 kHz. Furthermore, the L1 trigger sends the η and ϕ coordinates which caused the L1 trigger to fire, the so called *region of interest* (ROI), to the HLT.
The high level trigger uses information in the ROI at full granularity. It reduced the event rate to around 1 kHz with a decision time of 200 ms.
Several triggers are provided in a *trigger menu* based on the number of objects, amount of transverse momentum or missing transverse energy, and certain identification and isolation criteria. Fig. 3.8 shows different triggers and their rates for data taken in July 2016. The triggers which are used in this analysis are discussed

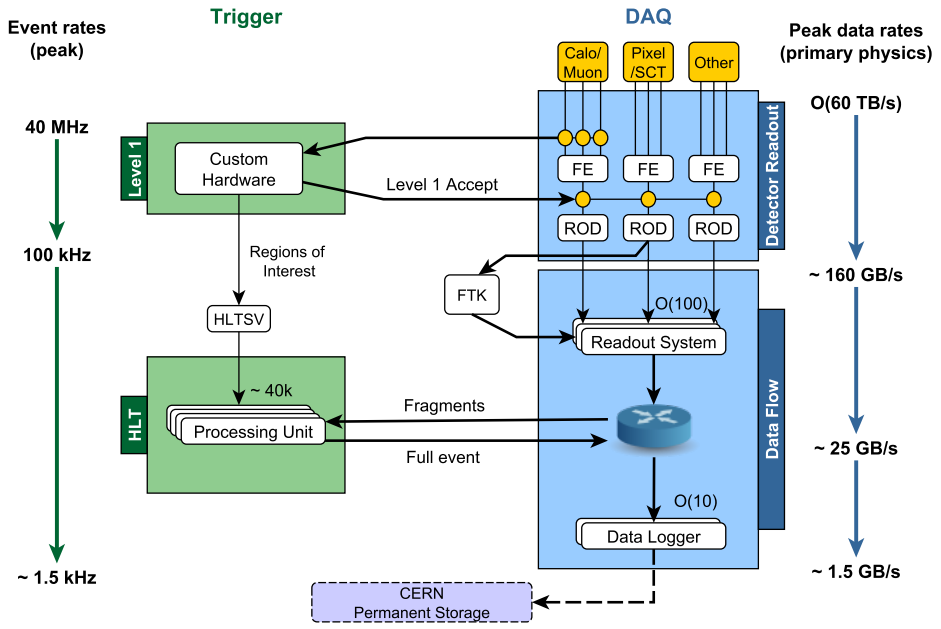


Figure 3.7.: The ATLAS trigger and data acquisition system for Run-2. [72]

in detail in Section 6.2.

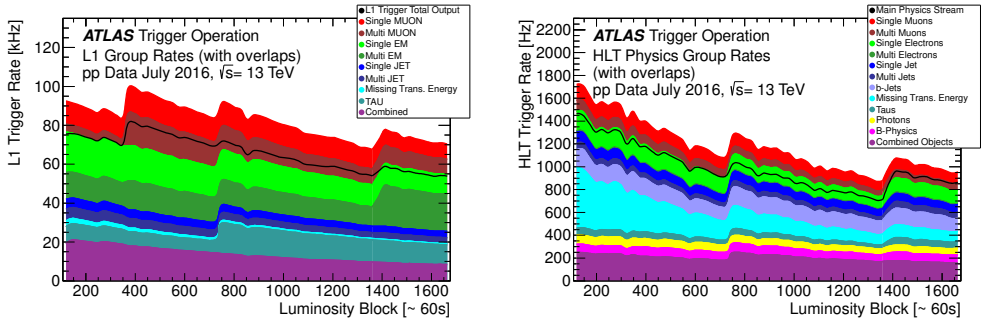


Figure 3.8.: Trigger menu and rates for the L1 trigger (left) and HLT (right) for data taken in July 2016 [73].

3.2.6. Data taking in 2015 and 2016

During 2015 and 2016 the ATLAS experiment recorded data from proton–proton collisions at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$, which corresponds to an integrated luminosity of $3.9\,\text{fb}^{-1}$ and $35.6\,\text{fb}^{-1}$, respectively. Not all data satisfies imposed data quality criteria, which are discussed in Section 6.3, therefore not all data can be used for physics analyses. This analysis uses data corresponding to

3.21 fb^{-1} and 32.86 fb^{-1} for 2015 and 2016, respectively. The uncertainty on the luminosity measurement is 2.1 % [74].
Because there are 10^{11} protons in each bunch it is likely that more than one interaction occurs per bunch crossing. This is called *in-time pile-up*. Due to the low time distance of 25 ns between each bunch crossing interactions which happen directly before or after the interaction of interest can also be recorded, since the read out time of the calorimeters is much slower. This is called *out-of-time pile-up*. The mean of total interactions per bunch crossing in data taken in 2015 and 2016 is 23.7 [70]. Fig. 3.9 shows a distribution of the mean pile-up for data taken in 2015 and 2016.

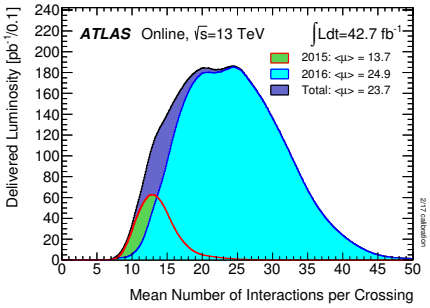


Figure 3.9.: Distribution of the mean number of interactions per bunch crossing (pile-up) weighted by luminosity for data taken in 2015 and 2016 [70].

4. Signal and Background Processes

This chapter discusses shortly the signal process which is considered in this analysis and gives an overview over the relevant background contributions. For a precise measurements it is important that the background processes are understood and modeled well. In Section 4.3 the used software to predict and simulate signal and background processes is discussed.

4.1. Signal Process

The Higgs boson has several production modes and decay channels, as discussed in Section 2.2. In this analysis the four main production mechanisms, namely the gluon–gluon fusion, vector-boson fusion, Higgs-Strahlung, and top-quark pair associated production are considered, which are explained in more detail in Section 2.2.1. The two dominant production modes are gluon–gluon fusion and vector-boson fusion. The decay of the Higgs boson into two τ-leptons, $H \to \tau\tau$ is analyzed, which has a branching ratio of 6.272 % for a mass of the Higgs boson of $m_H = 125\,\text{GeV}$. Since τ-leptons have a very short mean decay lifetime of $2.9 \times 10^{-13}\,\text{s}$ they cannot be detected directly. It is only possible to detect τ-leptons by reconstructing their decay products. A τ-lepton can decay either into leptons (electrons or muons) or hadrons (combinations of charged and neutral pions) via electroweak interactions, with a branching ratio of 35 % and 65 %, respectively [8]. Thus, the $H \to \tau\tau$ decay can be categorized into three subchannels, depending on the final state of the decaying τ-leptons. The focus of this analysis lies on the full-leptonic decay channel, $H \to \tau^+\tau^- \to \ell^+\ell^- 4\nu$, where both τ-leptons decay leptonically. This decay channel has a branching ratio of 12 %. The other two decay channels are the semi-leptonic and full-hadronic channel with one lepton and one hadron or two hadrons in the final state, respectively. The corresponding branching ratios are 46 % and 42 %.

4.2. Background Processes

At the LHC a lot of different interactions can happen during the collisions. Some processes have a similar detector signature as the signal process, the so-called *background*. There are two kinds of background processes, *reducible* and *irreducible* ones. Irreducible backgrounds have the same final state as the signal process (i.e. same number of leptons, jets, b-jets, etc.), which makes the separation very hard. For an accurate analysis, those backgrounds need to be modelled by simulations or estimated in a data driven way. But also processes with a different final

state topology can contribute to the background. Errors in the identification of physical objects or reconstruction of the missing transverse energy can lead to a signal-like detector signature, even if the real event topology is different from the signal process. These processes are called reducible. In the following sections all background processes are discussed, which are relevant to this analysis.

4.2.1. Z Boson Production in Association with Jets

One of the most important background processes is the production of a Z-boson or virtual photon, with a subsequent decay into τ-leptons or light leptons (electrons or muons). This background features both an irreducible and a reducible part. The $Z/\gamma^* \to \tau\tau \to \ell\ell + 4\nu$ background is irreducible, due to the same final state topology. If the Z-boson or virtual photon decay directly into electrons or muons, no transverse energy is produced. However, additional jets in the final state can lead to a misreconstruction of the missing transverse energy. Example Feynman diagrams of the production of a Z-boson with up to 2 additional jets in the final state are shown in Fig. 4.1.

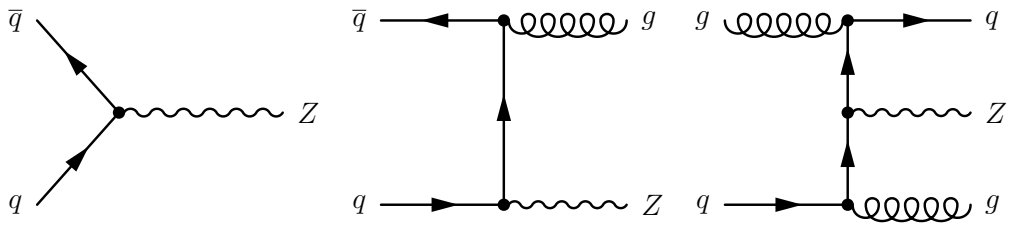


Figure 4.1.: Example Feynman diagrams for Z-boson production with up to two associated jets.

4.2.2. Diboson production

The production of WW-, WZ-, and ZZ-diboson pairs is combined in the diboson background. Here both W- and Z-bosons can decay either leptonically or hadronically. The most important contribution comes from WW-boson decays, $WW \to 2\ell 2\nu$, since they have the same final state as the signal process. Fig. 4.2 shows Feynman diagrams for the different diboson production mechanisms. Additionally, the decay of Higgs bosons into a pair of W-bosons is also considered as background.

4.2.3. Single Top-Quark and Top-Quark Pair Production

Another important background is the production of one or two top-quarks, whose decay is accompanied by large amounts of jets. Single top-quarks can be produced both in the s- and t-channel and in association with a W-boson, as shown in Fig. 4.3. The top-quarks decay in almost all cases into a b-quark and W-boson.

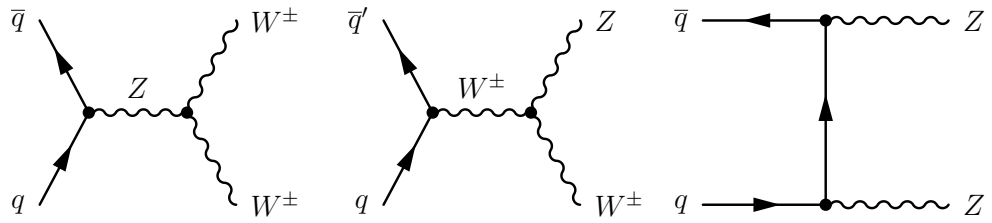


Figure 4.2.: Example Feynman diagrams for dominant diboson production modes.

Since the b-quark can also decay into a W-boson and a lighter quark, two leptons and missing transverse energy can be in the final state due to the decay of the W-bosons. Additionally, decays of B-hadrons can also provide a prompt lepton in the final state.

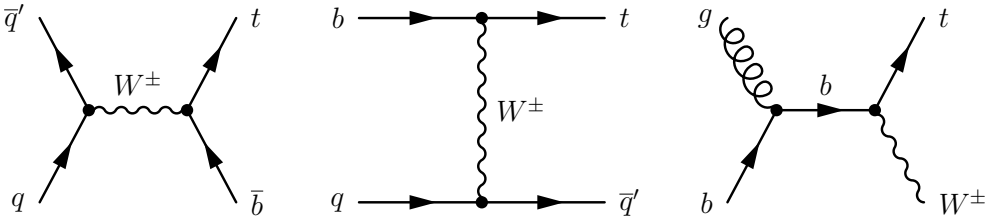


Figure 4.3.: Example Feynman diagrams for single top-quark production in the s-channel (left), t-channel (middle), and $tW^{\pm}$ production (right).

The production of top-quark pairs is however the more dominant part of this background. Top-quark pairs can be produced in processes with quarks and gluons in the initial state, which can be seen in Fig. 4.4. At the LHC the $gg \to t\bar{t}$ processed dominate, due to the high values of the gluon PDF in protons at low values of the momentum fraction x. The decay chain form top-quarks to leptons is described above.

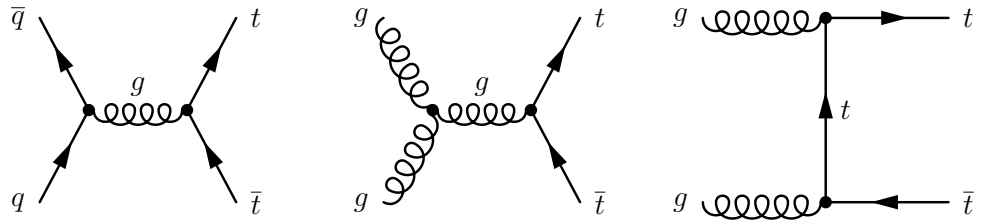


Figure 4.4.: Example Feynman diagrams for the production of top-quark pairs, $t\bar{t}$.

4.2.4. QCD Multi-Jet Production

Because protons are collided at the LHC, QCD interactions with outgoing quarks and gluons have a high cross-section. The quarks and gluons create jets due to

hadronization, which sometimes are misidentified as leptons. With the additional misreconstruction of missing transverse energy some events have a signal-like event topology. Example Feynman diagrams for QCD multi-jet processes are shown in Fig. 4.5.

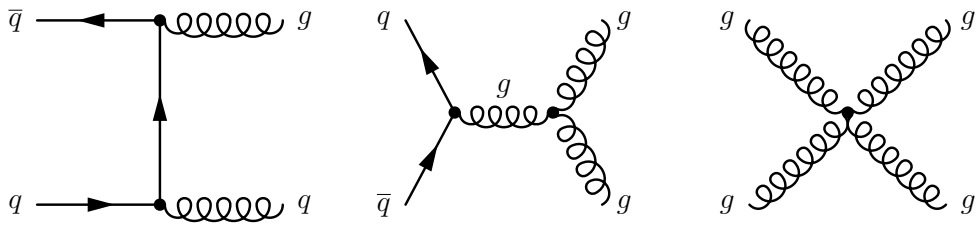


Figure 4.5.: Example Feynman diagrams for QCD multi-jet production.

4.3. Monte Carlo Simulations

All signal processes except for the production mode associated with a top-quark pair are modelled with Powheg-Box v2 [75] interfaced to Pythia8 [76]. For gluon–gluon fusion and vector-boson fusion the NNLOPS [77] PDF set is used for the matrix element and the AZNLO CTEQ6L1 [78] PDF tune for the modelling of non-perturbative effects, while for the Higgs-boson production associated with a vector boson the NNPDF3.0 and AZNLO CTEQ6L1 tune is used. For the simulation of $t\bar{t}H$ events aMC@NLO combined with Pythia8 is used. Here the PDFs are described by NNPDF3.0 [79]. All decay channels of the $H \to \tau\tau$ decay are included in the signal samples.
Events originating from the Z/γ^* and diboson background are generated by Sherpa 2.2.1 [80–84] with the NNPDF30NNLO [79] PDF tune. The electroweak contributions of Z/γ^* are calculated seperately with the same settings. To simulate events of the top-quark background a combination of Powheg and Pythia6 [85] is used with the CT10 PDF set and Perugia 2012 tune. The $H \to W^+W^-$ process is generated by Powheg and Pythia with the CT10 PDF set. The generators and cross-sections for all processes are listed in Fig. 4.6.
For all events the full response of the ATLAS detector is simulated [86] with the help of Geant4 [87]. Pile-up events are generated with Pythia8 and overlaid corresponding to the pile-up profile in data.

Figure 4.6.: Signal and background processes used in the $H \to \tau_{\text{lep}}\tau_{\text{lep}}$ analysis. The cross-section corresponds to a 2015 + 2016 dataset at $\sqrt{s} = 13\,\text{TeV}$. The generators and PDF sets which are used to predict the events are also listed.

Signal $m_H = 125\,\text{GeV}$	σ [pb] at $\sqrt{s} = 13\,\text{TeV}$	Order	Generator	PDF tune
ggF $H \to \tau\tau$	48.58 [46]	N3LO QCD + NLO EW	Powheg + Pythia8	NNLOPS + AZNLO CTEQ6L1
VBF $H \to \tau\tau$	3.782 [46]	NNLO QCD + NLO EW	Powheg + Pythia8	NNLOPS + AZNLO CTEQ6L1
WH $H \to \tau\tau$	1.373 [46]	NNLO QCD + NLO EW	Powheg + Pythia8	NNPDF30 + AZNLO CTEQ6L1
ZH $H \to \tau\tau$	0.8839 [46]	NNLO QCD + NLO EW	Powheg + Pythia8	NNPDF30 + AZNLO CTEQ6L1
ttH $H \to \tau\tau$	0.5071 [46]	NLO QCD + NLO EW	aMC@NLO + Pythia8	NNPDF3.0
Background				
$Z/\gamma^* \to \ell\ell$ ($\ell = e, \mu, \tau$)	1992 [88, 89]	NNLO	Sherpa 2.2.1	NNPDF30NNLO
Diboson	95.4 [90, 91]	NLO	Sherpa 2.2.1	NNPDF30NNLO
single t, s-channel	5.8 [92]	NNLO	Powheg + Pythia6	CT10 + Perugia 2012
single t, t-channel	87.8 [93]	NNLO	Powheg + Pythia6	CT10 + Perugia 2012
tW	22 [94]	NNLO	Powheg + Pythia6	CT10 + Perugia 2012
$t\bar{t}$	253 [95–99]	NNLO + NNLL	Powheg + Pythia6	CT10 + Perugia 2012
ggF $H \to W^+W^-$	48.58 [46]	N3LO QCD + NLO EW	Powheg + Pythia	CT10
VBF $H \to W^+W^-$	3.782 [46]	NNLO QCD + NLO EW	Powheg + Pythia	CT10

5. Object Selection

For any physics analysis in proton-proton collisions at the LHC the particles traversing the ATLAS detector immeres in magnetic fields need to be reconstructed and identified. For this information of the inner detector, calorimeters, and the muon spectrometer is used. The reconstruction is done both on data and simulation. Differences in reconstruction, identification, and trigger efficiencies are measured and correction factors (*scale factors*) are applied to the simulated events.
First, a brief overview of the reconstruction of tracks and vertices is given in Section 5.1, which is fundamental to the reconstruction of all other objects. Since this analysis is focusing on the $H \to \tau^+\tau^- \to \ell^+\ell^- 4\nu$ decay, the reconstruction and identification of electrons and muons is essential and described in Sections 5.2 and 5.3. Additional jets in the events are used to define a VBF and boosted topology and to suppress background (Section 6.4). Their reconstruction is explained in Section 5.4 as well as the identification of jets originating from B-hadrons. The identification of hadronically decaying τ-leptons (Section 5.5) is used to suppress additional background. Due to the four neutrinos in the final state the reconstruction of the missing transverse energy (Section 5.6) is also necessary. The chapter closes with Section 5.7, where the removal of the overlap between objects is explained.

5.1. Tracks and Vertices

Charged particle trajectories (*tracks*), as neutral particles do not leave tracks, are the fundamental ingredient for the reconstruction and identification of other physics objects, which are discussed in the following sections. The tracks are reconstructed in the inner detector by using various track reconstruction algorithms [100]. Information of the several ID sub-detector systems (IBL, Pixel, SCT, TRT) are taken into account. The transverse momentum and the sign of the charge of the track is calculated from the curvature of the track in the magnetic field. Quality criteria which are based on the number of hits in the sub-detectors depending on the transverse momentum p_{T} and pseudorapidity η are applied. The track reconstruction efficiencies range between 91 % and 63 % depending on the η of the track and different selection criteria [101].
Tracks can be used to reconstruct the primary and secondary interaction points (*vertices*). For this the tracks are extrapolated back to the interaction point to check for intersections between different tracks. Since multiple interactions are expected during one bunch crossing (*pile-up*) there are also multiple vertices, which are reconstructed. The vertex with the highest $\sum p^2_{\mathrm{T,track}}$ is chosen as the primary

vertex, which corresponds to the point where the interaction was the hardest. The efficiency to reconstruct a vertex depends on the number of tracks which are associated with the vertex. For two tracks the efficiency is 83 %, for three tracks 97 %, and for more than 4 tracks around 100 % [102].
Detailed information about tracking and vertexing in ATLAS for Run-2 of the LHC can be found in [101–103].

5.2. Electrons

Electrons can be identified by matching clusters of energy depositions in the electromagnetic calorimeter with extrapolations to the calorimeter of reconstructed tracks provided by the ID. To suppress background contributions from pile-up events or other objects like jets and hadronically decaying τ-leptons additional information of the ID and hadronic calorimeter are considered. The identification algorithm [104], a likelihood-based method, uses a multivariate analysis (MVA) technique to combine and evaluate all provided information. Different requirements on the likelihood discriminant[1] yield different operating points, labeled as *loose*, *medium*, and *tight*. They provide a different level of electron identification efficiency and background rejection. The reconstruction efficiencies range between 95 % for the loose working point and 80 % for the tight working point.
In this analysis the *loose* criterion is chosen. Additional requirements are $p_{\mathrm{T}} > 15\,\mathrm{GeV}$ and $|\eta| < 2.47$. The Pixel Detector and SCT can only provide information for reconstruction and identification in this η-range. Electrons within $1.37 < |\eta| < 1.52$ are excluded, because of the poor reconstruction and identification performance caused by the transition region between the barrel and end-cap calorimeters.
To increase the background rejection, isolation requirements are introduced by the following two discriminating variables to exlude electron candidates. The calorimetric isolation energy $E_{\mathrm{T}}^{\mathrm{cone0.2}}$ is defined as the sum of transverse energy deposited within $\Delta R = 0.2$ around the electron candidates. Corrections for electron energy leakage, pile-up, and the underlying event activity are applied. The sum of the transverse momentum of all tracks within $\Delta R = \min(0.2, 10\,\mathrm{GeV}/E_{\mathrm{T}})$ builds the track isolation $p_{\mathrm{T}}^{\mathrm{varcone0.2}}$. The tracks need to fulfill certain quality requirements and have to originate from the primary vertex. Based on different selection criteria on the quantities $E_{\mathrm{T}}^{\mathrm{cone0.2}}/E_{\mathrm{T}}$ and $p_{\mathrm{T}}^{\mathrm{varcone0.2}}/E_{\mathrm{T}}$ multiple operating points are constructed. This analysis uses the *gradient* isolation criterion with a targeted efficiency of $0.1143\,\% \times E_{\mathrm{T}}/\mathrm{GeV} + 92.14\,\%$ [104].
The efficiencies of the electron identification and isolation criteria are measured using a *tag-and-probe* technique in $Z \to e^+e^-$ and $J/\Psi \to e^+e^-$ events [104]. The combined reconstruction and identification efficiencies in $Z \to e^+e^-$ events as a function of E_{T} and η are shown in Fig. 5.1. To account and correct for differences in data and simulated events the efficiencies are calculated for both event types.

[1] The discriminant is defined as $\frac{\mathcal{L}_S}{\mathcal{L}_S+\mathcal{L}_B}$, where $\mathcal{L}_S$ and $\mathcal{L}_B$ denote the product of the signal and background probability density functions of the used variables.

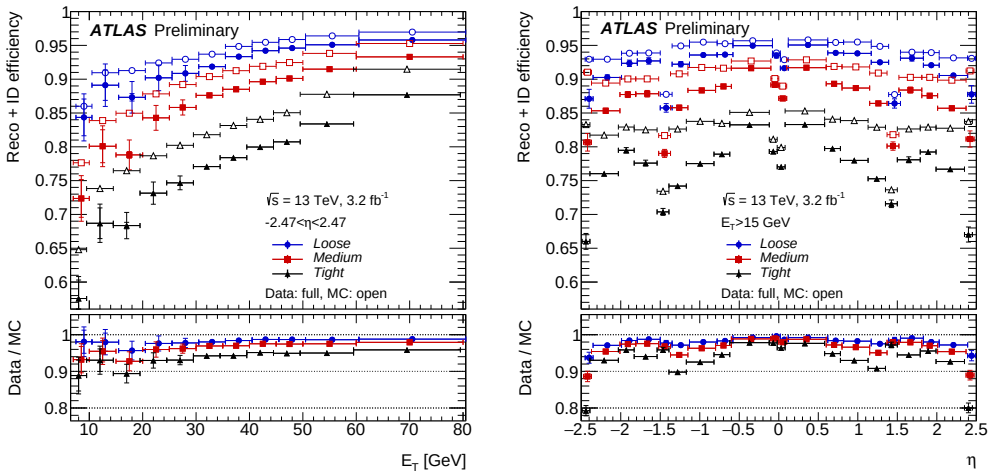


Figure 5.1.: Combined electron reconstruction and identification efficiencies for $Z \to e^+e^-$ events as a function of E_{T} (left) and η (right) for the *loose*, *medium*, and *tight* working points obtained for data and simulation. The inner error bars show the statistical uncertainty, the outer error bars combine the statistical and systematic uncertainties [104].

The ratio is then used to derive *scale factors*, which are applied to the simulated events in this analysis.

5.3. Muons

Muons are reconstructed by taking information of the inner detector (ID), calorimeter, and the muon spectrometer (MS) into account. Because muons traverse the detectors with minimum energy loss, they have a clear signature in the detectors and the discrimination between them and other physics objects like electrons and jets reaches a high accuracy.

First, muons are reconstructed independently in the ID and the MS. For the track reconstruction in the MS each of the muon chambers is searched for hit patterns, which are combined to track segments. The segments are then combined to muon track candidates. For each track candidate a global χ^2 fit is performed to the hits associated with the track. If a certain threshold is reached the track is accepted [105].

After the individual reconstruction four different algorithms are applied to combine the information of the different sub-detector systems. Combined (CB) muons have both a track in the ID and MS. The global track is calculated by a refit to both tracks. If there is only one local track segment in the MDT or CSC chambers and the track of the ID can be extrapolated to the MS the muon is classified as segment-tagged (ST). A track in the ID can be classified as calorimeter-tagged (CT) muon if the track can be matched to a energy deposition in the calorimeter, in the case that the energy deposition has the signature of a minimum-ionizing particle.

Extrapolated (ME) muons are reconstructed only from tracks in the MS with the additional requirement that the track needs to originate from the interaction point. To reduce background from mainly pion and kaon decays, different muon identification criteria are defined, called *loose*, *medium*, and *tight*. For *medium* muons only CB and ME tracks are used with some additional requirements on the number of hits in different layers and the *q/p significance*[2]. In this analysis *loose* muons are used with $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ and $|\eta| < 2.5$. This includes all *medium* muons as well as CT and ST muons, which are however restricted to $|\eta| < 0.1$. The reconstruction efficiency for tight muons is about 95 % and for medium muons 0.99 %.
Isolation requirements can further reduce background, because muons originating from heavy particles like W, Z, or Higgs bosons are often produced isolated in contrast to muons from b/c-flavoured hadrons. Two discriminating variables are introduced. The calorimetric isolation energy $E_{\mathrm{T}}^{\mathrm{topocone20}}$ is defined as the sum of transverse energy deposited within $\Delta R = 0.2$ around the muon candidates. Corrections for pile-up and the underlying event activity are applied. The sum of the transverse momentum of all tracks with $p_{\mathrm{T}} > 1\,\mathrm{GeV}$ within $\Delta R = \min(0.3, 10\,\mathrm{GeV}/p_{\mathrm{T}}^{\mu})$ excluding the muon candidate track is defined as the track isolation $p_{\mathrm{T}}^{\mathrm{varcone0.3}}$. Based on different selection criteria on the quantities $E_{\mathrm{T}}^{\mathrm{topocone20}}/p_{\mathrm{T}}^{\mu}$ and $p_{\mathrm{T}}^{\mathrm{varcone30}}/p_{\mathrm{T}}^{\mu}$ multiple operating points are constructed. This analysis uses the *gradient* isolation criterion which provides an efficiency of more than 90(99) % at 20(60) GeV [105].
The muon reconstruction and identification efficiencies are obtained with a *tag-and-probe* technique using $Z \to \mu^+\mu^-$ and $J/\Psi \to \mu^+\mu^-$ events [105]. Fig. 5.2 shows the reconstruction efficiencies for *medium* and *loose* muons. The efficiencies are calculated both in data and simulated events in order to derive *scale factors*, which are applied to the simulated events to correct deviations of the efficiencies between data and simulation.

[2]The *q/p significance* is the absolute value of the difference between charge and momentum measured in ID and MS divided by the sum of squares of the respective uncertainties.

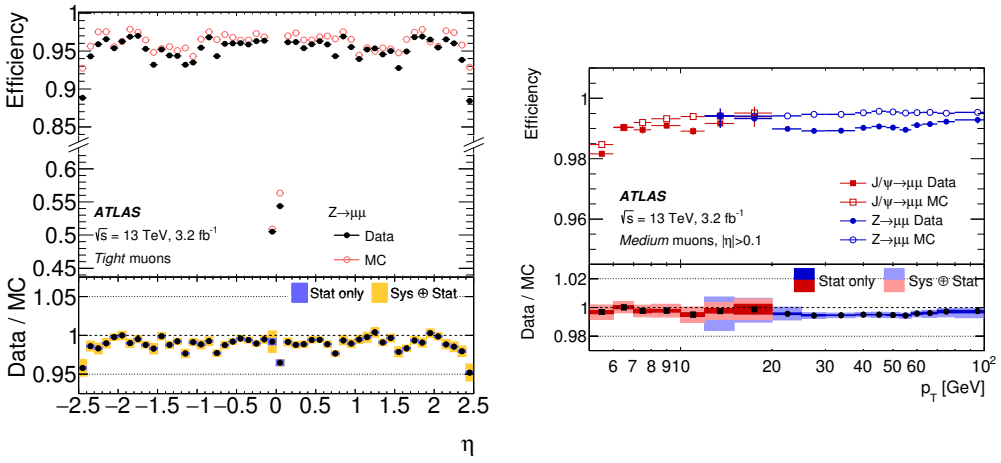


Figure 5.2.: Muon reconstruction efficiencies for *loose* and *medium* muons as a function of η measured in $Z \to \mu^+\mu^-$ events (left) and for *medium* muons as a function of p_{T} measured in $Z \to \mu^+\mu^-$ and $J/\Psi \to \mu^+\mu^-$ events (right) for data and simulation [105].

5.4. Jets

Particles with a color charge like quarks and gluons cannot exist in an unbound state, they form colorless states due to hadronization. Collimated bunches of hadrons are produced, which are called *jets*.

An algorithm which reconstructs jets should be insensitive to soft radiation (*infrared safety*) and splitting of the initial seed (*collinear safety*). There are two types of jet reconstruction algorithms, cone type and sequential clustering algorithms. Cone type algorithms use a geometrical cone around a jet axis to reconstruct the jet. In the past not all cone algorithms were infrared and collinear safe. Sequential cluster algorithms combine different object based on their energy and angular properties. They provide infrared and collinear safety by construction.

In this analysis the anti-k_t [106, 107] sequential clustering algorithm, based on energy clusters in the hadronic calorimeter, with a distance parameter of $R = 0.4$ is used for jet reconstruction. Not all jets are considered in the analysis, only jets with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and $|\eta| < 4.5$ are used.

The jet four-momenta undergo a series of corrections [108], to take several insufficiencies into account. First, the jet origin is corrected to point back to the primary vertex. Next excess energy due to pile-up is removed. Truth information of simulated dijet events is used to correct the jet energy scale (JES). Additional JES corrections are performed in the *global sequential calibration*, which uses calorimeter, muon spectrometer and track-based variables. Finally an *in-situ* correction in data is applied, using events in $Z + jets$, $\gamma + jets$, and dijet processes. These corrections raise a list of systematic uncertainties, which are discussed in Section 10.1.

Pile-up jets can be suppressed with the output of the jet vertex tagger (JVT) algorithm [109], which uses tracking and vertexing information to distinguish jets

from hard- and soft-scatter interactions. All jets in this analysis with $p_{\mathrm{T}} < 50\,\mathrm{GeV}$ and $|\eta| < 2.4$ are required to have $|\mathrm{JVT}| > 0.59$, where the output of the JVT algorithm is also labeled JVT. In the forward region a special algorithm for forward jets (fJVT) is used [110]. For this analysis jets with $p_{\mathrm{T}} < 50\,\mathrm{GeV}$ and $|\eta| > 2.5$ need to pass the fJVT algorithm with fJVT > 0.4.

Jets originating from b-quarks, also called b-jets, can be identified with *b-tagging* algorithms. These algorithms exploit the fact that b-flavoured hadrons (B-hadrons) have quite a long mean life time ($\tau \approx 1.5\,\mathrm{ps}$ [8]) compared to other hadrons. The decay creates a secondary vertex several millimeters away from the primary vertex due to time dilation.[3]. The secondary vertex is reconstructed with the tracks of the charged particles within a jet.
Since this analysis focuses on the gluon–gluon fusion and VBF production modes of the Higgs boson, jets produced from the decay of a b-quark are not expected. However, top quarks, which produce a large background in this analysis, almost always decay into b-quarks Therefore, identifying jets originating from b-hadrons gives the possibility to reduce the background produced by top quarks.
This analysis uses the multivariate-based b-tagging algorithm MV2c20 [111, 112] with a working point resulting in 85 % efficiency for b-jets in simulated $t\bar{t}$ events. The comparison to data is used to calculate tagging and mis-tagging correction factors. The requirements of $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $|\eta| < 2.4$ are additionally applied to the b-tagged jets.

5.5. Hadronically decaying τ-leptons

Tau leptons can either decay into leptons ($\tau \to \ell\nu_\tau\nu_\ell$, $\ell = e, \mu$) or hadrons ($\tau \to$ hadrons, denoted as τ_{had}). With a mass of 1.777 GeV and a proper decay length of 87 µm [8], the decay usually happens before the τ-lepton reaches the active part of the ID.
Leptonically decaying τ-leptons are identified as electrons ore muons, only hadronically decaying τ-leptons are reconstructed as τ-leptons. Most of the time the decay products are either one ore three charged pions or kaons, one ore more additional neutral pion can also be produced. Depending on the number of charged pions tracks, the event is called either *1-* or *3-pronged*. All visible decay products are denoted as $\tau_{\mathrm{had\text{–}vis}}$.
Since this analysis is focusing on the $H \to \tau^+\tau^- \to \ell^+\ell^- 4\nu$ decay, no hadronically decaying τ-leptons are expected. However a veto on $\tau_{\mathrm{had\text{–}vis}}$ candidates can be used to reduce background (see Section 6.3).
The reconstruction starts by selecting jets which are reconstructed by using the anti-k_t algorithm [106, 107] with a distance parameter of $\Delta R = 0.4$. Only jets with $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ and $|\eta| < 2.5$ are considered. $\tau_{\mathrm{had\text{–}vis}}$ candidates within the transition

[3] If a lifetime of $\tau = 1.5\,\mathrm{ps}$ and rest mass of $m_0 = 5.3\,\mathrm{GeV}$ are assumed (approximate values, taken from [8]), the B-hadron travels $c\tau' = c\tau\beta\gamma = c\tau\frac{E}{m_0} = 4.24\,\mathrm{mm}$ in the laboratory frame, if it has an energy of $E = 50\,\mathrm{GeV}$.

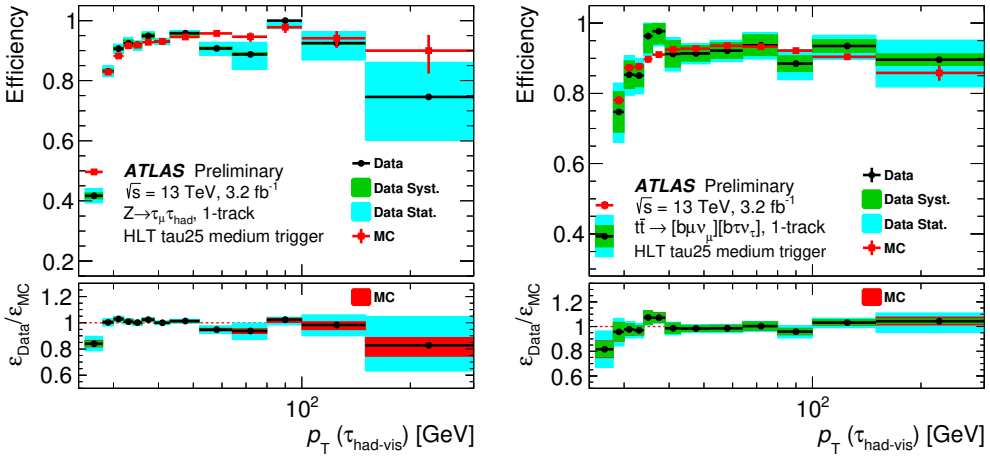


Figure 5.3.: Tau trigger efficiencies and scale factors ($\epsilon_{\text{Data}}/\epsilon_{\text{MC}}$) for 1-pronged hadronic τ decays, measured in $Z \to \tau\tau$ (left) and $t\bar{t}$ (right) events with the 2015 dataset for data and simulation [115].

region of the barrel and end-cap calorimeters ($1.37 < |\eta| < 1.52$) are discarded. A tau vertex is calculated from the associated tracks ($\Delta R < 0.2$, $p_{\text{T}} > 1\,\text{GeV}$) of the $\tau_{\text{had–vis}}$ candidates. The η-ϕ direction is determined with information from the calorimeters. The energy is obtained by a tau-specific calibration scheme [113].

Quark- and gluon-initiated jets have a more broader shower profile than jets caused by τ-leptons. This can be used to distinguish between the origin of the jets. For this a Boosted Decision Tree (BDT) based method is used. The BDT provides multiple working points labeled as *loose*, *medium*, and *tight* with a targeted efficiency of 0.6 (0.5), 0.55 (0.4), and 0.45 (0.35) for 1(3)-pronged $\tau_{\text{had–vis}}$ candidates, respectively.

In this analysis the $\tau_{\text{had–vis}}$ candidates need to pass the *loose* working point, with the additional requirements of $p_{\text{T}} > 20\,\text{GeV}$ and $|\eta| < 2.5$.

Since the BDT was trained to discriminate between quark- and gluon-initiated jets and jets from τ-leptons, it does not perform well with regard to discriminating between 1-pronged hadronic τ decays and electrons. A likelihood discriminator [113] is build to act as an *electron veto*, which uses the shower shape in the calorimeter and track information provided by the ID, including the TRT.

The reconstruction and identification efficiencies as well as the trigger efficiency of hadronically decaying τ-leptons is measured in $Z \to \tau_\mu \tau_{\text{had}}$ events with a *tag-and-probe* technique. The tau trigger efficiency can also be measured in $t\bar{t} \to \left[b\mu\nu_\mu\right]\left[b\tau\nu_\tau\right]$ events with a similar *tag-and-probe* method. To account for differences of efficiencies in data and simulation correction factors (scale factors) are derived and applied to simulated events. [114, 115] The efficiencies and scale factors for 1-pronged tau decays are illustrated in Fig. 5.3.

5.6. Missing Transverse Energy

In proton-proton collisions the exact momentum of the initial partons is not known. However, the assumption can be made that partons carry no transverse momentum [116]. Thus, the transverse momentum in the final state should also be zero, due to energy and momentum conservation. Any imbalance in the final state transverse momentum is known as the *missing transverse energy* and denoted as $E_{\mathrm{T}}^{\mathrm{miss}}$. Ideally it arises from weakly-interacting, stable particles produced in the collision. In the SM these particles are the neutrinos, but it may also be an indication of weakly-interacting exotic particles. Since the focus of this analysis is the full leptonic decay $H \to \tau^+\tau^- \to \ell^+\ell^- 4\nu$, a large $E_{\mathrm{T}}^{\mathrm{miss}}$ contribution is expected due to the four final state neutrinos. Additional *fake* $E_{\mathrm{T}}^{\mathrm{miss}}$ contributions can arise from pile-up or SM particles, which escape the detector without being detected, are badly reconstructed, or cannot be reconstructed at all. These contributions distort the real $E_{\mathrm{T}}^{\mathrm{miss}}$ and need to be corrected.
For the reconstruction of the missing transverse energy first the vectorial quantity $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$ is calculated with reconstructed and calibrated physics objects [117]

$$\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}} = \mathbf{E}_{\mathrm{T}}^{\mathrm{miss},e} + \mathbf{E}_{\mathrm{T}}^{\mathrm{miss},\gamma} + \mathbf{E}_{\mathrm{T}}^{\mathrm{miss},\tau} + \mathbf{E}_{\mathrm{T}}^{\mathrm{miss,jet}} + \mathbf{E}_{\mathrm{T}}^{\mathrm{miss,soft}} + \mathbf{E}_{\mathrm{T}}^{\mathrm{miss},\mu}\,, \tag{5.1}$$

with the missing transverse energy $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss,type}} = -\sum \mathbf{p}_{\mathrm{T}}^{\mathrm{type}}$ for each type of object (e: electrons, γ: photons, τ: τ-leptons, jet: jets, soft: soft objects, μ: muons). The individual contributions will be explained in the next paragraphs. Except for the soft and muon term all contributions originate from the calorimeters. Most of the time and also in this analysis the scalar missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$ is used,

$$E_{\mathrm{T}}^{\mathrm{miss}} = \left\|\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}\right\| = \sqrt{(E_x^{\mathrm{miss}})^2 + \left(E_y^{\mathrm{miss}}\right)^2}\,. \tag{5.2}$$

The objects which are taken for $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss},e}$, $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss},\mu}$, $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss},\tau}$, and $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss,jet}}$ are described in the sections above. The reconstruction of photons, which are needed to calculate $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss},\gamma}$, is covered in [118]. The contributions for $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss,soft}}$ originate from ID tracks associated with the primary vertex of the hard interaction, which are not used in the reconstruction of the other, high p_{T} objects. This is implemented in the Track Soft Term (TST) algorithm [117].
The performance of the $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction can be measured in $Z \to \mu^+\mu^-$ and $W^\pm \to e^\pm\nu$ events comparing data to simulation [119]. Those two processes provide events where either a fake or real dominated $E_{\mathrm{T}}^{\mathrm{miss}}$ contribution is expected. Fig. 5.4 shows the performance and resolution measured in $Z \to \mu^+\mu^-$ events.

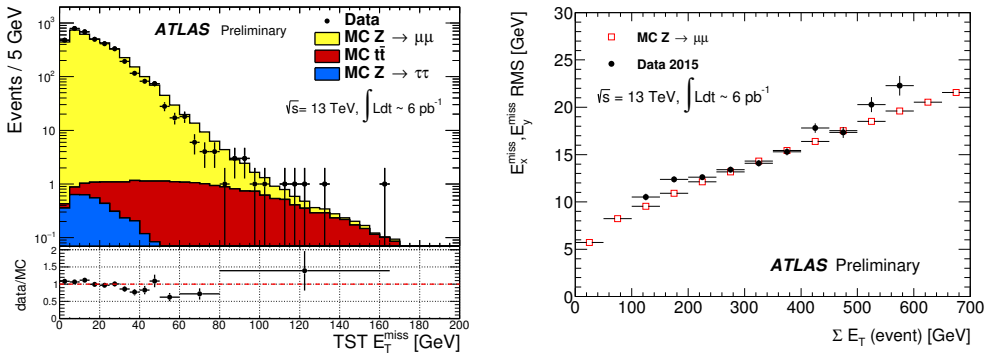


Figure 5.4.: Distributions of TST $E_{\mathrm{T}}^{\mathrm{miss}}$ (left) and the TST $E_x^{\mathrm{miss}}, E_y^{\mathrm{miss}}$ resolution (right) in $Z \to \mu^+\mu^-$ events. [119]

5.7. Overlap Removal

It is possible that a detector signature of one particle can pass the reconstruction and identification requirements of multiple objects. To remove this kind of ambiguity from the analysis an overlap removal is applied. If two ore more objects are reconstructed within a certain distance ΔR in the $(\eta - \phi)$ plane, only one object is kept. First jets are removed, then hadronic τ-leptons, and finally electrons. Muons are ordered highest and are not removed. Depending on the object combination, a different ΔR threshold is used:

1. hadronic τ-leptons
   - remove jets within $\Delta R = 0.2$
2. electrons
   - remove jets within $\Delta R = 0.4$
   - remove hadronic τ-leptons within $\Delta R = 0.2$
3. muons
   - remove jets within $\Delta R = 0.4$
   - remove hadronic τ-leptons within $\Delta R = 0.2$
   - remove electrons within $\Delta R = 0.2$

6. Event Selection

The $H \to \tau^+\tau^- \to \ell^+\ell^- 4\nu$ decay has a detector signature of exactly two light leptons (e or μ) of opposite electric charge and some amount of missing transverse energy due to the neutrinos. The lepton combinations can be either two electrons (e^+e^-), two muons ($\mu^+\mu^-$), or one electron and one muon ($e^\pm\mu^\mp$). The decay channels are labeled as ee and $\mu\mu$ for same flavour (SF) combinations and $e\mu$ or μe for different flavour (DF) combinations. The leptons and jets are each arranged according to their transverse momentum in descending order. They are labeled with integer numbers, starting with 1. The lepton or jet with the highest transverse momentum is called *leading*, the next one is denoted as *subleading*. If a quantity uses both leading and subleading lepton or jet, it will be indicated by a $\ell\ell$ or jj label, respectively.

This chapter gives an overview of the selection criteria applied to select signal candidate events. The requirements are applied to both data and simulated events. Since the mass reconstruction in the $H \to \tau^+\tau^- \to \ell^+\ell^- 4\nu$ decay channel is non-trivial due to the multiple neutrinos, advanced mass reconstruction methods are introduced in Section 6.1. Triggers (Section 6.2) are used to select events with two final state leptons. After a common preselection (Section 6.3) further selection criteria are applied to select events which fall in the VBF and boosted topology (Section 6.4). This allows to separate the Higgs bosons which are produced via the VBF and ggF mechanism and to optimize the sensitivity of the measurement.

The selection criteria introduced in this chapter belongs to the cut-based analysis (CBA), which is the baseline for the multivariate analysis (MVA) developed in the context of this thesis. The MVA is presented in Chapter 9, where the slightly different event selection is discussed.

6.1. Invariant mass reconstruction

The invariant mass of the Higgs boson candidates (i.e. the invariant mass of the di-τ system) cannot be calculated without ambiguity from only the two leptons and $E_{\mathrm{T}}^{\mathrm{miss}}$, since there are four neutrinos in the final state of the $H \to \tau_{\mathrm{lep}}\tau_{\mathrm{lep}}$ decay. A correct and precise reconstruction is needed, because the di-τ mass can be used to discriminate between signal and background processes. In the next sections two approaches are introduced, the collinear approximation and the missing mass calculator.

6.1.1. Collinear approximation

For the collinear approximation [120,121] it is assumed that the $E_{\mathrm{T}}^{\mathrm{miss}}$ originates only from the neutrinos in the $H \to \tau\tau$ decay and that each τ-lepton is emitted in the same direction as its corresponding visible decay product (the lepton). The second assumption is called collinearity. This is a valid assumption, since $m_H/2 \gg m_\tau$, which leads to highly boosted τ-leptons.
With those two assumptions the invariant mass m_{coll} of the di-τ system can be calculated with

$$m_{\mathrm{coll}} = \frac{m_{\mathrm{vis}}}{\sqrt{x_1 x_2}}\,, \tag{6.1}$$

where m_{vis} is the mass of the visible decay products of the τ-lepton decay. The momentum fraction, which each visible decay product holds in comparison to the decaying τ-lepton, is denoted as $x_{1,2}$,

$$p_{\mathrm{T}}^{\ell_i} = x_i p_{\mathrm{T}}^{\tau_i}\,, \qquad i = 1, 2\,. \tag{6.2}$$

They can be calculated with

$$x_{1,2} = \frac{p_x^{\ell_1} p_y^{\ell_2} - p_y^{\ell_1} p_x^{\ell_2}}{p_x^{\ell_1} p_y^{\ell_2} \pm E_x^{\mathrm{miss}} p_y^{\ell_{2,1}} - p_y^{\ell_1} p_x^{\ell_2} \mp E_y^{\mathrm{miss}} p_x^{\ell_{2,1}}}\,. \tag{6.3}$$

The collinear approximation works well when the di-τ system is boosted and the approximations are valid. However, if the two taus are back-to-back ($\Delta\phi(\tau_1, \tau_2) = \pi$), the missing transverse energy due to the neutrinos cancels partially and the equation system which results in Eq. (6.3) cannot be solved anymore. This can be prevented by a requirement on the ϕ difference of the visible decay products, $\Delta\phi_{\ell\ell}$, or a direct cut on the momentum fractions, which discards events with unphysical solutions ($x < 0$ or $x > 1$).

6.1.2. Missing mass calculator

If the assumption of collinearity of the decay products of the τ-leptons is not made, there is no unique solution for the invariant mass of the di-τ system. By using on-shell conditions for the τ-lepton masses and assuming that all $E_{\mathrm{T}}^{\mathrm{miss}}$ originates from the τ-lepton decays, the following set of equations can be constructed.

$$\begin{aligned}
E_x^{\mathrm{miss}} &= p^{\mathrm{miss}_1} \sin\theta_{\mathrm{miss}_1} \cos\phi_{\mathrm{miss}_1} + p^{\mathrm{miss}_2} \sin\theta_{\mathrm{miss}_2} \cos\phi_{\mathrm{miss}_2} \\
E_y^{\mathrm{miss}} &= p^{\mathrm{miss}_1} \sin\theta_{\mathrm{miss}_1} \sin\phi_{\mathrm{miss}_1} + p^{\mathrm{miss}_2} \sin\theta_{\mathrm{miss}_2} \sin\phi_{\mathrm{miss}_2} \\
m_{\tau_1}^2 &= m_{\mathrm{miss}_1}^2 + m_{\ell_1}^2 + 2\sqrt{p^{\ell_1} + m_{\ell_1}^2}\sqrt{p^{\mathrm{miss}_1} + m_{\mathrm{miss}_1}^2} \\
&\quad - 2p^{\ell_1} p^{\mathrm{miss}_1} \cos\left(\theta_{\ell_1} - \theta_{\mathrm{miss}_1}\right) \\
m_{\tau_2}^2 &= m_{\mathrm{miss}_2}^2 + m_{\ell_2}^2 + 2\sqrt{p^{\ell_2} + m_{\ell_2}^2}\sqrt{p^{\mathrm{miss}_2} + m_{\mathrm{miss}_2}^2} \\
&\quad - 2p^{\ell_2} p^{\mathrm{miss}_2} \cos\left(\theta_{\ell_2} - \theta_{\mathrm{miss}_2}\right)
\end{aligned} \tag{6.4}$$

The known variables are the components of the missing transverse energy (E_x^{miss} and E_y^{miss}) and the momenta and invariant masses of the visible decay products of

the τ-leptons (ℓ_1 and ℓ_2). Unknown are the momenta, masses, and angles (ϕ and θ) of the two neutrino systems, which are composed of two neutrinos each. These quantities are labeled with miss_1 and miss_2 depending on the associated τ-lepton. Since this system of equations is underconstrained there is no single solution. However, the different solutions have different probabilities to occur due to the matrix element of the τ-decay. The missing mass calculator (MMC) algorithm [122] scans over four of the unknown variables ($m_{\text{miss}_{1,2}}$ and $\Phi_{\text{miss}_{1,2}}$), solves the system of equations and assigns each solution the corresponding probability obtained by calculating the matrix element of the τ-decay with the solution-specific kinematics. The resolution of the E_x^{miss} and E_y^{miss} variables is also included in the scan, because the MMC algorithm is affected by the resolution of $E_{\text{T}}^{\text{miss}}$. The algorithm can fail to converge if the $E_{\text{T}}^{\text{miss}}$ is badly reconstructed. To obtain only one solution either the most probable solution can be chosen or an average over all solutions, weighted by the corresponding probability, can be calculated. In this analysis the former approach is chosen.

6.2. Trigger

Triggers are used for the initial decision if an event will be further analyzed or discarded. For this analysis events with two leptons (ee, $e\mu$[1], $\mu\mu$) in the final state are selected. A general overview of the ATLAS trigger system was given in Section 3.2.5. Both single-lepton and dilepton triggers are used. They differ between the 2015 and 2016 data taking period due to the increase in luminosity.

Table 6.1.: Single-lepton triggers and offline p_{T} thresholds for the 2015 and 2016 data set used in the $H \to \tau_{\text{lep}}\tau_{\text{lep}}$ analysis.

Lepton flavour	Year	Trigger	offline p_{T} Threshold

Electron	2015	`HLT_e24_lhmedium_L1EM20VH`
		`HLT_e60_lhmedium`
		`HLT_e120_lhloose`
			$p_{\text{T}} > 25\,\text{GeV}$

	2016	`HLT_e26_lhtight_nod0_ivarloose`
		`HLT_e60_lhmedium_nod0`
		`HLT_e140_lhloose_nod0`
			$p_{\text{T}} > 27\,\text{GeV}$

Muon	2015	`HLT_mu20_iloose_L1MU15` `HLT_mu50`	$p_{\text{T}} > 21\,\text{GeV}$
	2016	`HLT_mu26_ivarmedium` `HLT_mu50`	$p_{\text{T}} > 28\,\text{GeV}$

The triggers for the $H \to \tau_{\text{lep}}\tau_{\text{lep}}$ analysis are listed in Tables 6.1 and 6.2. Multiple triggers for the same object (or object combination) and data taking period are

[1] In this section $e\mu$ indicates all events with one electron and one muon, the differentiation between $e\mu$ and μe is not made here.

Table 6.2.: Dilepton triggers and offline p_{T} thresholds for the 2015 and 2016 data set used in the $H \to \tau_{\mathrm{lep}}\tau_{\mathrm{lep}}$ analysis.

Channel	Year	Trigger	offline p_{T} Threshold
ee	2015	`HLT_2e12_lhloose_L12EM10VH`	$p_{\mathrm{T}}^{e_1,e_2} > 15\,\mathrm{GeV}$
	2016	`HLT_2e17_lhvloose_nod0`	$p_{\mathrm{T}}^{e_1,e_2} > 18\,\mathrm{GeV}$
$e\mu$	2015	`HLT_e17_loose_mu14`	$p_{\mathrm{T}}^{e} > 18\,\mathrm{GeV}, p_{\mathrm{T}}^{\mu} > 15\,\mathrm{GeV}$
	2016	`HLT_e17_lhloose_nod0_mu14`	$p_{\mathrm{T}}^{e} > 18\,\mathrm{GeV}, p_{\mathrm{T}}^{\mu} > 15\,\mathrm{GeV}$
$\mu\mu$	2015	`HLT_mu18_mu8noL1`	$p_{\mathrm{T}}^{\mu_1} > 19\,\mathrm{GeV}, p_{\mathrm{T}}^{\mu_2} > 10\,\mathrm{GeV}$
	2016	`HLT_mu22_mu8noL1`	$p_{\mathrm{T}}^{\mu_1} > 24\,\mathrm{GeV}, p_{\mathrm{T}}^{\mu_2} > 9\,\mathrm{GeV}$

combined with a logical 'or'. The overlap of single-lepton and dilepton triggers is avoided to prevent difficulties with the trigger efficiencies by selecting the triggers based on the transverse momentum of the lepton. Lower p_{T} thresholds are used for dilepton triggers and higher ones for single-lepton triggers as illustrated in Fig. 6.1.

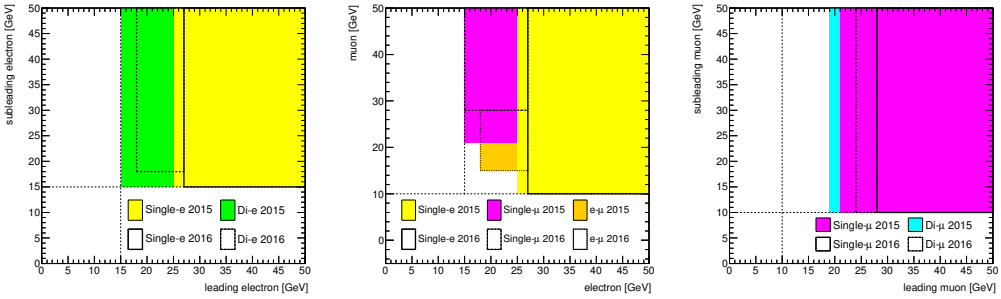


Figure 6.1.: Transverse momentum criteria to select either single-lepton or dilepton triggers in order to avoid overlap. The plots correspond to ee (left), $e\mu$ (middle), and $\mu\mu$ (right) final states [123].

Trigger names are composed of a series of acronyms and abbreviations chained together, directly or with underscores, which define the trigger type and the imposed requirements on the objects to trigger. For this analysis all trigger names start with `HLT`, indicating that the software-based High Level Trigger is used. Electrons and muons are denoted as `e` and `mu`, respectively, followed by a number defining the transverse momentum threshold in GeV, which the object has to fulfill. A preceding number corresponds to multiple objects with the same requirements. Identification and isolation criteria as introduced in Sections 5.2 and 5.3 can be imposed on the triggered objects, which is decoded as `lhID` and `iISO`, where `ID` and `ISO` specify the corresponding working points. If there is no requirement on the distance of the observed tracks to the primary vertex the term `nod0` is included.
HLTs can be seeded by L1 triggers, which is indicated by appending the L1 trigger name to to the HLT name. All L1 trigger names start with `L1`. The next part indicates which subdetector fired the L1 trigger. The electromagnetic calorimeter

is referred to as `EM` and the muon spectrometer as `MU`. Next, the transverse energy or transverse momentum threshold in units of GeV is denoted, depending on the subdetector. The E_{T} thresholds can vary slightly as a function of η, which is denoted with `V`. An additional veto on energy depositions in the hadronic calorimeter can be applied, which is indicated by `H`.
To account for differences in the trigger efficiencies, correction factors are calculated by comparing the efficiencies between data and simulation [124, 125]. An additional offline p_{T} requirement is introduced to ensure that the trigger efficiencies are in the plateau region. The thresholds are 1 GeV to 3 GeV higher than the trigger p_{T} thresholds. They can be found in Tables 6.1 and 6.2.

6.3. Preselection

First, all events are selected which pass the triggers as described in the section above. Before applying kinematic cuts to suppress background contributions a series of requirements to ensure data quality are applied.
Data events are discarded if they are not included in the good run list, which contains the runs where all subdetector systems have been in full operational mode. The total integrated luminosity of all the good runs in 2015 and 2016 is $\int \mathcal{L}\,\mathrm{d}t = 36.1\,\mathrm{fb}^{-1}$. Additionally, at least one reconstructed vertex consistent with the IP is required. This rejects events from cosmic rays and beam-halo effects. Furthermore, badly reconstructed events are removed.
Now basic preselection requirements (*cuts*) are applied to select the decay topology of the $H \to \tau_{\mathrm{lep}}\tau_{\mathrm{lep}}$ decay. The cuts are enumerated for future references. In Figs. 6.2 and 6.3 distributions of the signal and different backgrounds is shown for a observable before the cut on this observable is performed. Normalization factors as defined in Section 7.2 are applied. Their values are 1,06 for single and pair production of top quarks, and 1.19 and 1.07 for $Z \to \ell\ell$ and $Z \to \tau\tau$ production, respectively. The signal is scaled by factor 50. Underflow and overflow bins are included in the first and last bin, respectively. Only statistical uncertainties are contained in the error band.

(1) Number of leptons
Exactly two leptons, either two electrons, one electron and on muon, or two muons, with the reconstruction and identification criteria defined in Sections 5.2 and 5.3 are required.

(2) Lepton identification and isolation criteria
The electrons and muons need to pass the *medium* identification and *gradient* isolation criteria.

(3) Hadronic tau veto
To ensure orthogonality to the $H \to \tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ and $H \to \tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channels all events with one or more hadronic τ-leptons obeying the *medium* criterion are vetoed.

(4) Trigger
The event is associated with the trigger which selected it by testing which trigger is passed by the event. Since the trigger efficiency is not the same for data and simulated events, correction factors are derived by comparing the trigger efficiency in data and simulation. Those correction factors need to be applied to the simulated events. However, different triggers have different efficiencies and correction factors, so it is important that the correction factor of the right trigger is applied.

(5) Trigger matching
The objects which caused the event to pass the trigger have to match with the reconstructed leptons.

(6) Opposite sign
The electric charge of the two leptons has to be opposite.

(7) Dilepton mass
The mass of the dilepton system is restricted to $30\,\text{GeV} < m_{\ell\ell} < 75(100)\,\text{GeV}$ for SF (DF) events. For SF events this helps to reduce the Drell-Yan $Z/\gamma^* \to \ell\ell$ background, since the signal distribution peaks at around 50 GeV and the $Z \to \ell\ell$ background has its maximum at the mass of the Z boson, $m_Z = 91\,\text{GeV}$, as shown in Fig. 6.2a. The $\gamma^* \to \ell\ell$ background is reduced by the lower cut, because it rises for low values of $m_{\ell\ell}$. The main discrimination for DF events is against the top background, which is the dominating background outside of the range spanned by the cut. The distribution of $m_{\ell\ell}$ for DF events is displayed in Fig. 6.2b.

(8) Jet momentum
At least one jet with $p_\text{T} > 40\,\text{GeV}$ is required. This helps to select both the VBF and boosted category (Section 6.4). The transverse momentum distribution of the leading jet, $p_\text{T}^{\text{jet}_1}$, before Cut 8 is shown in Fig. 6.2c. Mainly $Z \to \ell\ell$ and $Z \to \tau\tau$ events are rejected by this cut.

(9) Missing transverse energy
Since the final state includes four neutrinos, a cut on $E_\text{T}^\text{miss} > 20\,\text{GeV}$ is applied for DF events. The $Z/\gamma^* \to \ell\ell$ background can be suppressed for SF by increasing this cut to $E_\text{T}^\text{miss} > 55\,\text{GeV}$. The distributions for E_T^miss before this cut are shwon in Figs. 6.2d and 6.2e for SF and DF events, respectively.

(10) Object based missing transverse energy (HPTO)
The missing transverse energy is calculated from only high-p_T objects (HPTO). Only the two decay leptons and all jets with a transverse momentum of $p_\text{T} > 30\,\text{GeV}$ are used.

$$\mathbf{E}_\text{T}^{\text{miss,HPTO}} = -\mathbf{p}_\text{T}^{\ell_0} - \mathbf{p}_\text{T}^{\ell_1} - \sum_{\substack{\text{jets} \\ p_\text{T} > 30\,\text{GeV}}} \mathbf{p}_\text{T}^\text{jet} \tag{6.5}$$

A HPTO missing transverse energy of $E_{\mathrm{T}}^{\mathrm{miss,HPTO}} > 55\,\mathrm{GeV}$ for SF events is required to further reject $Z/\gamma^* \to \ell\ell$ background, which can be seen in Fig. 6.2f.

(11) Momentum fraction

The momentum fractions carried by the visible decay products of the τ-lepton decay are restricted to $0.1 < x_{1,2} < 1.0$. They are calculated in the collinear mass approximation (Section 6.1.1). This rejects background events where the assumption of collinearity of the τ-lepton is not met, which leads to unphysical solutions. Figs. 6.3a and 6.3b show the distributions of x_1 and x_2 before Cut 11.

(12) Angular difference in η

This cut limits the η difference between the two leptons to $|\Delta\eta_{\ell\ell}| < 1.5$ to suppress background from single top and top-pair production. The $|\Delta\eta_{\ell\ell}|$ distribution before Cut 12 is shown in Fig. 6.3c.

(13) Angular difference in ΔR

In order to suppress the $Z \to \ell\ell$ and top background, the ΔR separation between the two leptons is required to be $\Delta R_{\ell\ell} < 2$. Fig. 6.3d shows the $\Delta R_{\ell\ell}$ distribution before Cut 16.

(14) Collinear mass

To ensure orthogonality with the $H \to WW$ analysis, the collinear mass (Section 6.1.1) is limited to $m_{\mathrm{coll}} > m_Z - 25\,\mathrm{GeV}$. Here, a value of $m_Z = 91.1876\,\mathrm{GeV}$ is used for the mass of the Z boson. The m_{coll} distribution before Cut 13 is shown in Fig. 6.3e.

(15) MMC mass

Events where the MMC mass reconstruction algorithm (Section 6.1.2) did not converge are discarded.

(16) b-jet veto

Events which contain b-jets (Section 5.4) with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ are vetoed. This helps to reduce the single-top and $t\bar{t}$ background.

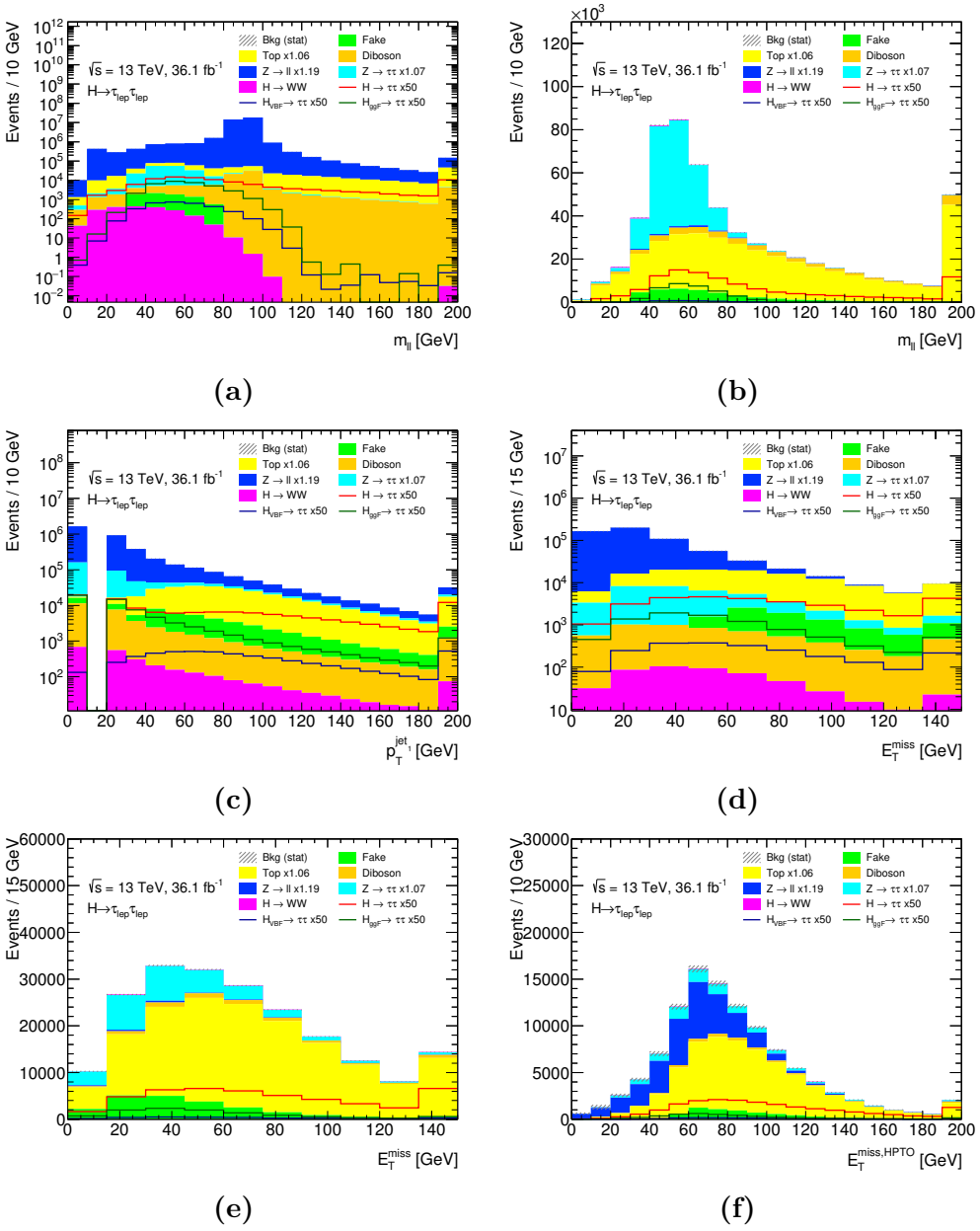


Figure 6.2.: Distribution of several observables which are used in the preselection before the corresponding cut is applied. The following observables are shown: $m_{\ell\ell}$ for SF events after Cut 6 (a), $m_{\ell\ell}$ for DF events after Cut 6 (b), $p_{\mathrm{T}}^{\mathrm{jet}_1}$ after Cut 7 (c), $E_{\mathrm{T}}^{\mathrm{miss}}$ for SF events after Cut 8 (d), $E_{\mathrm{T}}^{\mathrm{miss}}$ for DF events after Cut 8 (e), and $E_{\mathrm{T}}^{\mathrm{miss,HPTO}}$ for SF events after Cut 9 (e). The signal and background distributions are normalized to their theory cross-section and luminosity. Additional normalization factors are applied on the top-quark, $Z \to \ell\ell$, and $Z \to \tau\tau$ background. The signal is scaled by a factor of 50. Underflow and overflow bins are included in the first and last bin, respectively. Only statistical uncertainties are included in the error band.

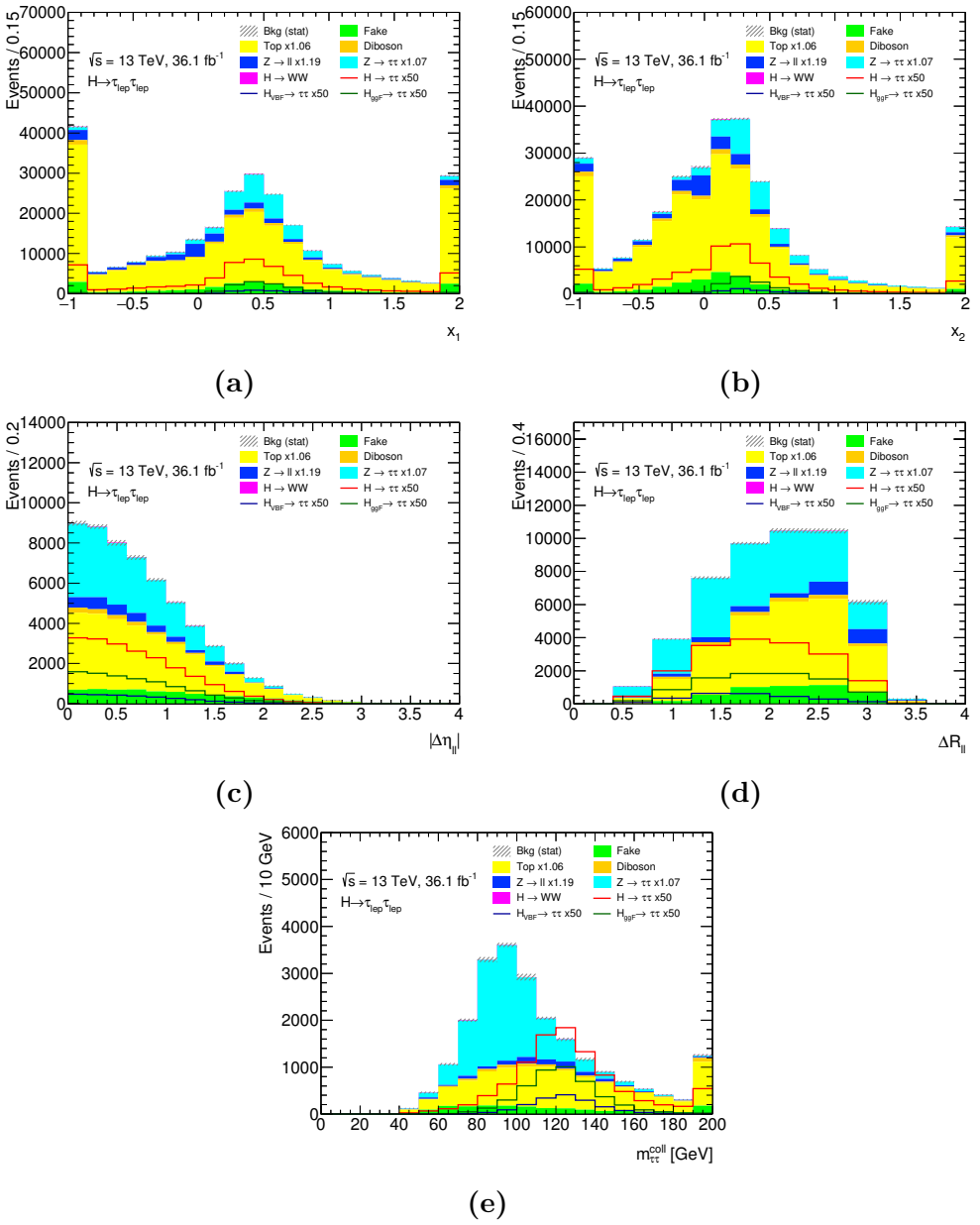


Figure 6.3.: Distribution of several observables which are used in the preselection before the corresponding cut is applied. The following observables are shown: x_1 after Cut 10, x_2 after Cut 10, $|\Delta\eta_{\ell\ell}|$ after Cut 11, $\Delta R_{\ell\ell}$ after Cut 12, and m_{coll} after Cut 13. The signal and background distributions are normalized to their theory cross-section and luminosity. Additional normalization factors are applied on the top-quark, $Z \to \ell\ell$, and $Z \to \tau\tau$ background. The signal is scaled by a factor of 50. Underflow and overflow bins are included in the first and last bin, respectively. Only statistical uncertainties are included in the error band.

6.4. Categorization

Since the goal of this analysis is to measure the coupling strength of the Higgs boson in different production modes, dedicated categories for the VBF and gluon–gluon fusion production mode are defined by exploiting production-mode specific event topologies. They are referred to as the *VBF* and *boosted category*, respectively. The signal-to-background ratio and background composition is different in those categories. Due to the splitting the overall sensitivity of the measurement is enhanced. Further subcategories formed in both cases to improve the sensitivity even more.

6.4.1. VBF category

In the VBF production mode two vector bosons are used to produce the Higgs boson. The vector bosons originate from two partons, which produce two jets with high transverse momentum in the forward and backward region of the detector. Small jet activity is expected between the two VBF jets, because there is no color flow between the initial partons. The following cuts are applied to select this topology:

- Subleading jet momentum (Cut 1V)
  A second jet is required with at least $p_{\mathrm{T}} > 30\,\mathrm{GeV}$, because the VBF topology has two jets. Fig. 6.5a shows the distribution of the transverse momentum of the second jet before this cut.

- Opposite hemispheres (Cut 2V)
  The two leading jets most likely occupy different hemispheres of the detector due to the VBF topology. This can be achieved by applying the $\eta_{\mathrm{jet}_1} \cdot \eta_{\mathrm{jet}_2} < 0$ requirement. The distribution of $\eta_{\mathrm{jet}_1} \cdot \eta_{\mathrm{jet}_2}$ before Cut 2V is shown in Fig. 6.5b.

- Angular separation of two leading jets (Cut 3V)
  The separation in η between the two VBF jets is expected to be large, as shown in Fig. 6.5c. Therefore, a cut of $\left|\Delta\eta_{\mathrm{jj}}\right| > 3$ is applied.

- Lepton candidate centrality (Cut 4V)
  The η of the selected leptons must lie in between the two jets.

- Invariant mass of the dijet system (Cut 5V)
  The dijet system is required to have an invariant mass of $m_{\mathrm{jj}} > 400\,\mathrm{GeV}$. The distribution of m_{jj} before Cut 5V is shown in Fig. 6.5d.

Furthermore, the VBF category is split into a *high* and *low* VBF category, with requirements on the transverse momentum of the di-τ system, $p_{\mathrm{T}}^{\tau\tau} > 100\,\mathrm{GeV}$ and $p_{\mathrm{T}}^{\tau\tau} < 100\,\mathrm{GeV}$, respectively. The transverse momentum of the di-τ system is calculated from the transverse momenta of the visible decay products of the τ-leptons and the missing transverse energy. This helps to increase the sensitivity, provided that the statistics in the subcategories are still high enough.

6.4.2. Boosted category

In contrast to the VBF production mode, the gluon–gluon fusion production mode has no outgoing partons at tree-level. However, higher order QCD corrections can produce one or more jets, which leads to a high transverse momentum of the Higgs boson. An example Feynman diagram for such a process is displayed in Fig. 6.4.

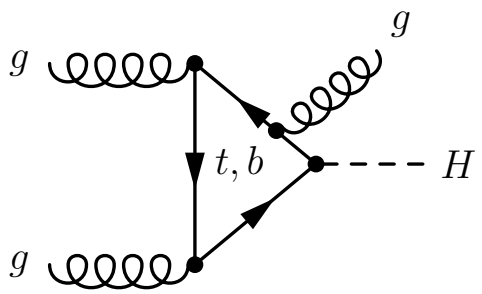


Figure 6.4.: Production of a Higgs boson via gluon–gluon fusion with an additional jet.

The selection criteria for the boosted category are as follows:

- Veto on VBF selection (Cut 1B) The events have to pass the preselection, but not the VBF selection.
- Higgs boson transverse momentum (Cut 2B) The transverse momentum of the di-τ system is required to be $p_{\mathrm{T}}^{\tau\tau} > 100\,\mathrm{GeV}$, since the goal is to select the boosted topology. A distribution of $p_{\mathrm{T}}^{\tau\tau}$ before Cut 2B is shown in Fig. 6.5e.

Similar to the VBF category, also the boosted category is divided into two subcategories. All events which pass the requirements $p_{\mathrm{T}}^{\tau\tau} > 140\,\mathrm{GeV}$ and $\Delta R_{\ell\ell} < 1.5$ are sorted into the *high-boosted* category. All other events which do not pass these criteria are filled in the *low-boosted* category.

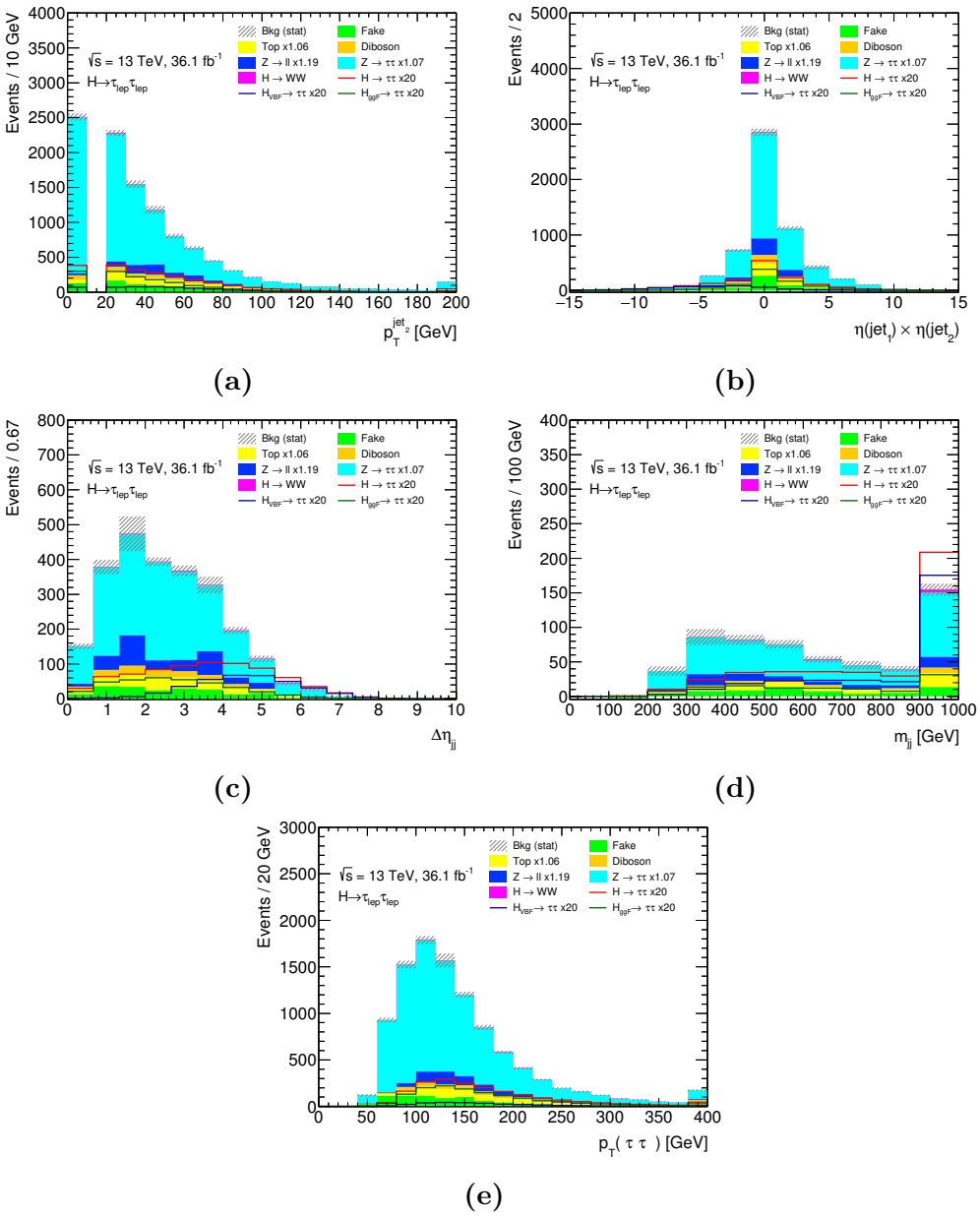


Figure 6.5.: Distribution of several variables which are used in the categorization before the corresponding cut is applied. The following distributions are shown: $p_{\mathrm{T}}^{\mathrm{jet}_2}$ after Cut 16 (a, the first bin contains all events with no second jet), $\eta_{\mathrm{jet}_1} \cdot \eta_{\mathrm{jet}_2}$ after Cut 1V, $\Delta\eta_{\mathrm{jj}}$ after Cut 2V, m_{jj} after Cut 4V, and $p_{\mathrm{T}}^{\tau\tau}$ after Cut 1B. The signal and background distributions are normalized to their theory cross-section and luminosity. Additional normalization factors are applied on the top-quark, $Z \to \ell\ell$, and $Z \to \tau\tau$ background. The signal is scaled by a factor of 20 for visibility. Underflow and overflow bins are included in the first and last bin, respectively. Only statistical uncertainties are included in the error band.

6.5. Event yields

The event yields for the different signal and background processes after the preselection and in the inclusive VBF and boosted categories are shown in Table 6.3. After the preselection the fraction of signal events is 1.05 %. For events from the ggF and VBF production mode the fractions are 0.67 % and 0.27 %, respectively. In the VBF category the proportion of the VBF $H \to \tau\tau$ process increases to 2.65 %, but also the ggF production mode contributes now 0.93 %. The overall contribution signal events to the total event count is 3.63 %.
It was also possible to enhance the signal fraction in the boosted category. The total signal contribution increases to 1.09 %. The proportions of the ggF and VBF production mode are 0.79 % and 0.12 %. The fraction of VBF events is slightly worse than after the preselection, because most VBF events fell in the VBF category. A large fraction of background events was rejected in the VBF category compared to the boosted category. This can be explained by the distinct event topology of a VBF event, which makes it easier to separate signal and background events.
The most dominant background is the $Z/\gamma^* \to \tau\tau$ background with 56 % (72 %) contribution for the VBF (boosted) category, followed by the $Z/\gamma^* \to \ell\ell$ and top-quark background with a fraction of 12 % (7 %) and 11 % (7 %), respectively. The QCD multijet ("fake") background has a contribution of 11 % in the VBF category and 0.7 % in the boosted category.

Table 6.3.: Event yields for the different signal and background processes after the preselection and in the inclusive VBF and boosted categories with a combined 2016 and 2016 dataset of 36.1 fb^{-1}. Normalization factors are applied on the top-quark, $Z \to \ell\ell$, and $Z \to \tau\tau$ background. Only statistical uncertainties are shown.

Process	Preselection	VBF category	Boosted category
ggF $H \to \tau\tau$	74.17 ± 0.55	5.18 ± 0.15	60.71 ± 0.49
VBF $H \to \tau\tau$	28.97 ± 0.15	14.77 ± 0.10	12.88 ± 0.10
WH $H \to \tau\tau$	5.48 ± 0.20	0.10 ± 0.02	4.94 ± 0.19
ZH $H \to \tau\tau$	2.91 ± 0.12	0.04 ± 0.01	2.67 ± 0.11
ttH $H \to \tau\tau$	2.32 ± 0.20	0.12 ± 0.05	2.12 ± 0.19
Fakes	841.34 ± 31.75	59.06 ± 8.83	529.25 ± 26.28
Top	684.23 ± 11.79	60.42 ± 3.36	540.07 ± 10.66
Diboson	438.00 ± 5.43	21.55 ± 1.10	368.10 ± 4.51
$Z/\gamma^* \to \ell\ell$	684.86 ± 85.54	64.73 ± 12.96	567.98 ± 79.12
$Z/\gamma^* \to \tau\tau$	7992.32 ± 62.85	322.03 ± 10.52	5541.34 ± 50.81
$H \to WW$	46.36 ± 1.20	8.66 ± 0.36	32.63 ± 1.07
Total signal	113.85 ± 0.65	20.21 ± 0.19	83.33 ± 0.58
Total background	10 687.11 ± 111.56	536.45 ± 19.21	7579.37 ± 98.33

7. Background Estimation

The background contributions as described in Section 4.2 play an important role in the analysis of the $H \to \tau_{\text{lep}}\tau_{\text{lep}}$ process. Most background contributions are modeled by simulated events. However, data-driven background estimation techniques allow a better control of the modeling of the simulated events while simultaneously reducing systematic uncertainties.
For this analysis the background contribution of events with jets misidentified as leptons is estimated in a data-driven way. Additionally, the normalization of simulated events from the $Z/\gamma^* \to \ell\ell$, $Z/\gamma^* \to \tau\tau$, and top-quark background is determined in so-called *control regions*. The normalizations are treated as nuisance parameters in the fit determining the signal strength, and are constrained from the event yields in the control regions.

7.1. Background Estimation of Events with Misidentified Leptons

Sometimes other objects like jets are misidentified as leptons. This happens most of the time in events which originate from QCD multi-jet production, W-boson production in association with jets and semi-leptonic decay of top-quark pairs. This background is called *fake background*. In contrast, real leptons can be found in processes with a so-called *prompt* lepton, for example the leptonic decay of a τ-lepton or a massive vector boson.
The estimation of the fake background is based on a control region where the isolation criterion on the subleading lepton is inverted. Additionally, the identification criterion is loosened from *medium* to *loose* and some requirements for the preselection are modified. For events with different-flavour (DF) final-state leptons, the invariant mass of the dilepton system has to fulfill $30\,\text{GeV} < m_{\ell\ell} < 150\,\text{GeV}$. A further requirement is $n_{\text{jets},40} \geq 1$, and if there is no jet in the event, the transverse momenta of the two leading leptons are restricted to $p_{\text{T}}^{\ell_1} > 35\,\text{GeV}$ and $p_{\text{T}}^{\ell_2} > 15\,\text{GeV}$.[1] For same flavour events also $n_{\text{jets},40} \geq 1$ is used instead of the normal jet counting.
Events from backgrounds without fake leptons (diboson, $Z \to \ell\ell$, $Z \to \tau\tau$, $H \to WW$, and leptonically decaying top quarks) are subtracted from data in this control region. The remaining data events form the *fake distribution*. However, this distribution is not correctly normalized. The normalization is obtained by fitting

[1] The notation $n_{\text{jets},40}$ indicates, that for the jet counting only jets with $p_{\text{T}} > 40\,\text{GeV}$ and $|\eta| < 4.5$ are used.

the p_{T} distribution of the subleading lepton in the control region with a template fit to the same distribution in the signal region (nominal region).
To correct for discrepancies between the nominal and fake region, correction factors are calculated in two different regions and then applied to the fake region. For this events with two same-sign (SS) leptons are used, which are obtained by inverting the requirement of the opposite sign (OS) of the two leptons. The correction factors f_{cor} are calculated by

$$f_{\mathrm{cor}} = \frac{N^{\mathrm{SS}}_{\mathrm{nom}}}{N^{\mathrm{SS}}_{\mathrm{fake}}}\,, \tag{7.1}$$

where N denotes the number of events in the corresponding region. Now, the correction factors are applied to the OS fake region,

$$N^{\mathrm{OS}}_{\mathrm{nom}} = f_{\mathrm{cor}} N^{\mathrm{OS}}_{\mathrm{fake}}\,. \tag{7.2}$$

The correction factors are applied to the distributions of the following observables: $p^{\ell_1}_{\mathrm{T}}, p^{\ell_2}_{\mathrm{T}}, \Delta\phi(\ell_1, E^{\mathrm{miss}}_{\mathrm{T}}), \Delta\phi(\ell_2, E^{\mathrm{miss}}_{\mathrm{T}}), n_{\mathrm{jets},40}, pt^{\tau\tau}, p^{\tau}_{\mathrm{T}}/p^{\ell_1}_{\mathrm{T}}, \Delta R_{\ell\ell}, m^{\ell_1}_{\mathrm{T}}$, and $m^{\ell_2}_{\mathrm{T}}$. The assumption was made, that the correction factors are not changed when transferring them from the SS to the OS region.

7.2. Normalization of the $Z \to \ell\ell$, $Z \to \tau\tau$, and top-quark produced backgrounds

The other important backgrounds of this analysis, i.e. the $Z \to \ell\ell$, $Z \to \tau\tau$, and top-quark produced background, are not estimated with a data-driven technique. Their shape is taken from simulations, but the normalization can be obtained from data events in control regions. Similarly to the estimation of the fake lepton background, the control regions are here also defined by inverting one requirement of the event selection. For each background contribution a separate control region is needed, which should be enriched with events from the corresponding background. The normalization of the fake background is calculated when constructing this background, therefore no further normalization is required. The other background contributions are not considered.
The $Z \to \ell\ell$ control region is defined by changing the requirement of $30\,\mathrm{GeV} < m_{\ell\ell} < 75\,\mathrm{GeV}$ to $80\,\mathrm{GeV} < m_{\ell\ell} < 100\,\mathrm{GeV}$, in order to select the Z-peak. Additionally, only same flavour events are considered. Events where the MMC algorithm failed to converge are also used, because only the event yields are used and not the shape information of the m_{MMC} distribution.
The control region for the background originating from top-quarks are formed by requiring at least one jet which is b-tagged in each event. Furthermore, events where the MMC algorithm failed to converge are also included.
For the $Z \to \tau\tau$ background there is no dedicated control region. Instead the signal region is used, but only events with $m_{\mathrm{coll}} < 100\,\mathrm{GeV}$ are used to avoid high signal contributions.

To obtain the normalization factors a 3×3 matrix $\mathbb{N}$ is constructed, which contains the number of events for each of those three backgrounds ($Z \to \ell\ell$, $Z \to \tau\tau$, top-quark production) in each control region. Additionally, a vector $\mathbf{N}^{\text{Data}}$ contains the number of data events in each control region. Contributions from other backgrounds which are not normalized need to be subtracted. Now, the vector of normalization factors, **NF**, can be calculated by multiplying the inverted matrix with the data vector,

$$\mathbf{NF} = \mathbb{N}^{-1}\mathbf{N}^{\text{Data}}\,. \tag{7.3}$$

Uncertainties on the normalization factors can be obtained by varying the entries of the $\mathbb{N}$ matrix within their uncertainties multiple times and then inverting each of the matrices.
These normalization factors are also called pre-fit normalization factors, since they are used to compare distributions before the statistical analysis is carried out, in which the normalizations are incorporated as a part of the full fit model.
The normalization factors can be in principle calculated after each requirement listed in Sections 6.3 and 6.4. For the analysis presented in Chapter 6 the normalization factors are calculated after the preselection. The event yields used to calculate the normalization factors are shown in Table 7.1. The results are shown in Table 7.2.

Table 7.1.: Event yields for the different background processes in the control regions after the preselection witha combined 2015 and 2016 dataset of $36.1\,\text{fb}^{-1}$.

Process	Top-quark CR	$Z \to \ell\ell$ CR	$Z \to \tau\tau$ CR
Signal	337.9 ± 2.5	1.46 ± 0.09	113.85 ± 0.65
Top-quark	$55\,582 \pm 104$	48.7 ± 2.9	645 ± 11
$Z \to \ell\ell$	799 ± 87	2413 ± 79	576 ± 72
$Z \to \tau\tau$	3599 ± 57	9.67 ± 1.44	7469 ± 59
Other	5633 ± 64	118.77 ± 2.32	1326 ± 32
Data	69863	3037	10630

Table 7.2.: Pre-fit normalization factors for the multivariate analysis.

Selection	$Z \to \ell\ell$	$Z \to \tau\tau$	Top
Preselection	1.19 ± 0.05	1.06 ± 0.02	1.07 ± 0.02

8. Boosted Decision Trees

As shown in Section 6.5 only a small fraction of all events which are analyzed originate from the signal process. All other events are produced by background processes. Since the goal is to extract information from the signal, one challenge of this analysis is to reduce background contributions as much as possible.
An established approach to separate signal and background is to restrict observables to a given range. The observables and thresholds (cuts) are usually selected and optimized by hand, which can be a tedious task. However, it is very easy to implement in the analysis and the effect of each threshold is easily comprehensible. Multiple observables are used to reject different types of background contributions. Such an approach is described in Chapter 6.
Another way to achieve separation between signal and background is with the help of machine learning (ML) algorithms. The most common ML algorithms used in high energy physics are *boosted decision trees* (BDTs) and neural networks. In this thesis an approach to use BDTs for the analysis of the $H \to \tau_{\text{lep}}\tau_{\text{lep}}$ process in the 2015 + 2016 dataset is developed. This chapter focuses on the general and theoretical aspects of boosted decision trees, while the application of BDTs in the analysis is discussed in Chapter 9. For this the ROOT [126] library TMVA [127] is used.
First, decision trees are introduced in Section 8.2, since they are the foundation of BDTs. Then, the concept of boosting is explained in Section 8.3 and different boosting algorithms are presented. Finally, a summary of BDT parameters is given in Section 8.4.
The contents of this chapter are based on Refs. [127, 128].

8.1. Introduction to Machine Learning

Machine learning has two different applications. It can be used to assign class labels to the data points (*classification*) or to predict the repsonse to input data encoded in a continuous variable via *regression*. For this analysis classification is used, since the goal is to split the events in signal and background. In general, there can be an arbitrary amount of output classes, but to simplify the discussion this chapter focuses only on two-class problems in the aspect of splitting signal and background events.
Both types of machine learning rely on the same principle. First, a model has to be trained with data, the so-called *training set*, where the correct result is already known. The data used can either come from measurements or simulation. After the model is trained it can be applied to another set of data. It is important that

the data which was used for training the model is not used in the real measurement, because the model would perform better on the training set than on a independent data set. This issue is discussed in more detail in Section 9.2.
It can happen that a model performs much worse on a independent data set than on the training set. This effect is known as *overtraining*. In this case the model was trained too much and is sensitive to statistical fluctuations in the training set. Overtraining can be avoided if the correct parameters in the training of the model are chosen.

8.2. Decision Trees

A decision tree has a binary-tree-like structure as shown in Fig. 8.1. The root node contains all signal and background events. All events are sorted into two subnodes, based on a certain observable (input variable) and threshold. Now the splitting procedure is recursively applied again on each of the subnodes, until a stopping criterion is met. The final notes are labeled as signal- or background-like, depending on the majority of their contents. A signal-like node results in an output of $+1$ of the decision tree, a background-like node yields -1. After the decision tree is built it can be applied to data events, where it classifies the data events as either signal- or background-like.

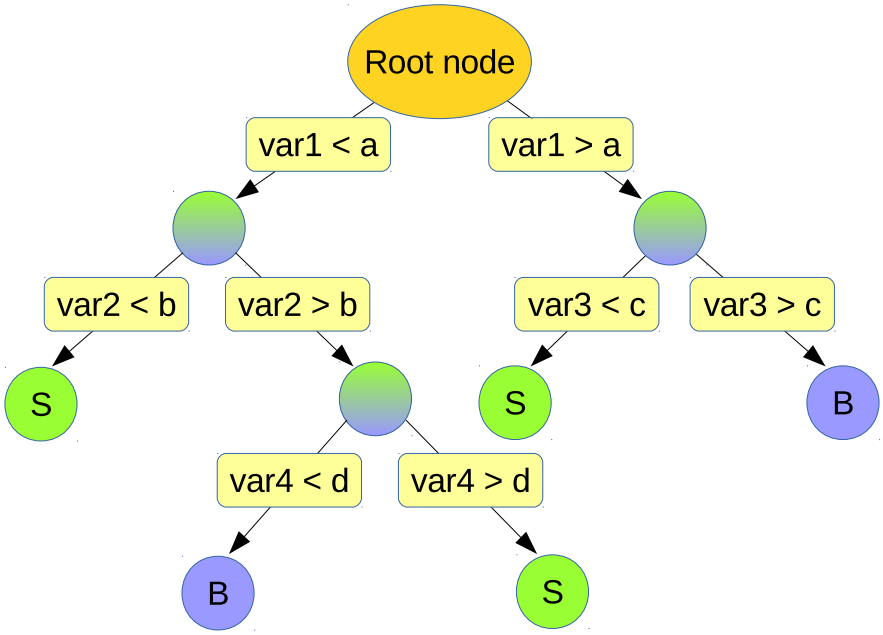


Figure 8.1.: Exemplary structure of a decision tree with a depth of 3. Signal- and background-like nodes are labeled with a S and B.

A decision tree can be interpreted as a function $G(x_i)$, which takes a vector of

input variables of one event x_i, with the following output:

$$G(x_i) = \begin{cases} +1 & \text{signal-like} \\ -1 & \text{background-like} \end{cases} . \tag{8.1}$$

If some signal and background events have similar properties, the decision tree cannot reach a perfect performance. Some signal events are classified as background events and vice versa. This is called *misclassification*. A good decision tree has only a small number of misclassified events.

8.2.1. Growing a decision tree

The process of deciding which observable and threshold is used to split each node is called *growing* or *building* a decision tree. The goal of the splitting is to improve the pureness in either signal or background events in the subnodes with respect to the parent node.
The signal purity p of a node is defined as

$$p = \frac{s}{s+b}, \tag{8.2}$$

where s is the number of signal events and b the number of background events. A node with only signal events has the purity of one, a node containing only background events has the purity of zero.
Based on the purity of the node a separation criterion Q is calculated. The separation criterion is a function which reaches its maximum when the node is fully mixed, i.e. for $p = 0.5$. Below some examples of different separation criteria are given.

- Gini index: $Q = p(1-p)$. The maximum value is $Q(0.5) = 0.25$.
- Cross entropy: $Q = -p\ln(p) - (1-p)\ln(1-p)$. The maximum value is $Q(0.5) \approx 0.69$
- Misclassification error: $Q = 1 - \max(p, 1-p)$. The maximum value is $Q(0.5) = 0.5$.

To increase the pureness of a node the increase of the separation criterion has to be optimized. The change of the separation criterion for a selected input variable v and threshold t is the difference between the sum of the new separation criteria on the two new subnotes, weighted by the relative fraction of events, and the separation criterion of the parent node,

$$\Delta Q(v,t) = \left| \sum_{i=1,2} \frac{n_i}{n_{\text{parent}}} Q_i(v,t) - Q_{\text{parent}} \right| . \tag{8.3}$$

Here n_i is the number of events in each subnode, n_{parent} the number of events in the parent node, Q_i the value of the separation criterion of each subnode, and Q_{parent}

the value of the separation criterion in the parent node. The maximum value of ΔQ is selected by considering all given input variables. For each variable a scan over the its range in the training set is performed. Since there is a finite amount of events in each node it is possible to change the thresholds in such a way that after each change only one event is sorted differently. However, this would need large amounts of computational power and is therefore not feasible in most applications. Thus, a common approach is to only test a certain amount of thresholds, which are distributed over the whole range of the variable.
In principle the splitting can go on until each node contains only one event, but this would lead to an overtrained decision tree. This can be prevented by *pruning* the decision tree, a process where nodes with low numbers of events are combined. Another way to stop growing the decision tree is to impose some cancellation criteria, so that the fine splitting does not happen in the first place. For example the maximal depth of the tree can be limited or nodes are only split again if they contain a certain number of events.

8.2.2. Comparison to Selection Cuts

Using decision trees has some advantages over the standard approach (denoted as cut-based analysis, CBA) of sequentially requiring thresholds on different observables. Usually all events which do not satisfy one requirement are discarded immediately. In decision trees all events are kept. It can happen that a signal event is first sorted into a background-like node, but later on the node is split again and the event falls into a signal-like node. Therefore, decision trees should provide an improved classification performance. This can be visualized by looking at the space which is spanned by the input variables. The cut-based approach selects only one hypercube in the feature-space[1]. Decision trees can select multiple hypercubes in this space.
Additionally, decision trees (and also other machine-learning models) are also sensitive to correlations between the input variables, which enables the decision trees to better classify events.

8.2.3. Disadvantages

Decision trees are not a perfect solution, they also have their drawbacks. In machine learning generally all models have a certain dependence on statistical fluctuations of the training set. If the training set is split into two sets and two individual decision trees are trained, they should in principle have a similar structure. However, due to statistical fluctuations one variable and threshold could be selected differently, which could lead to a completely different structure of one decision tree.
Also, decision trees are *weak classifiers*, i.e. they provide only a performance slightly better than random guessing [128].

[1] In machine learning, the input variables are called *features*.

However, there are concepts which provide substantial improvements in both of those areas. One method is the co-called *boosting*, which is introduced in the next section.

8.3. Boosting

The idea of boosting is to combine many weak classifiers into a powerful *committee*. This increases the performance compared to using only one decision tree. It is a very powerful way of improving the separation power of decision trees while simultaneously reducing the probability of overtraining.
The general concept of boosting is as follows. First one decision tree is trained on the initial training set. Weights are assigned to each event based on the output of the decision tree. If an event is misclassified a higher weight is assigned to it. Now, a new decision tree is trained on the weighted training set. Because previously misclassified events have a higher weight they have more impact on the current training step than corrctly classified events. This procedure is repeated multiple times, until a maximum number of iterations is reached. The final output is a combination of the output of all decision trees which were trained. This combination is known as the *boosted decision tree*.
There are different algorithms which implement this general boosting strategy. In the next sections first the *AdaBoost* algorithm is introduced, which is a simple boosting algorithm. It can be shown, that AdaBoost is a specific case of a general concept, which leads then to the second boosting algorithm which is discussed, the *gradient boost*.

8.3.1. AdaBoost

The AdaBoost algorithm [129] was developed initially in 1997 by Yoav Freund and Robert Schapire. Until today it is one the most popular boosting algorithms.
Let x be a set of N vectors of observables (input events) and y the corresponding output values. Single events are referred to as x_i and y_i with $i \in \{1, \ldots, N\}$, respectively. The classifier, in this case the decision tree, is denoted as $G(x_i)$, with $G(x_i) \in \{-1, 1\}$. The output values are either 1 for a signal event or -1 for a background event.
In the beginning each event is assigned the same weight w_i, where

$$w_i = \frac{1}{N} \,. \tag{8.4}$$

For each boosting step m the classifier is trained with the training events using the weights w_i. After the training the misclassification error, err_m, is calculated, which is the weighted average of the fraction of misclassified events,

$$\mathrm{err}_m = \frac{\sum_{i_1}^{N} w_i I(y_i \neq G_m(x_i))}{\sum_{i_1}^{N} w_i} \,. \tag{8.5}$$

The misclassification error is zero if all events are correctly classified and one if no event is correctly classified. Here the indicator function I is used, which is defined as follows:

$$I(cond.) = \begin{cases} 1 & \text{if } cond. \text{ is true} \\ 0 & \text{else} \end{cases} . \tag{8.6}$$

The boost weight α_m is calculated,

$$\alpha_m = \log \frac{1 - \text{err}_m}{\text{err}_m} . \tag{8.7}$$

It regulates how much the weights are altered for each misclassified event in the next iteration. The weights which are used in the next boosting step $m + 1$ are set in the following way:

$$w_i^{m+1} = w_i^m \exp \left[\alpha_m I \left(y_i \neq G_m(x_i) \right) \right] . \tag{8.8}$$

Only the weights of the events which are misclassified are altered. After the maximum number of boosting iterations, M, the final output $G(x)$ of the boosted decision tree can be calculated. The output of all decision trees weighted by the boost weight of this step is summed up,

$$G(x) = \frac{1}{M} \sum_{m=1}^{M} \alpha_m G_m(x) . \tag{8.9}$$

In contrast to the discrete output of single decision trees the combined output can have a real value between -1 and 1.
The learning rate can be adjusted by replacing the boost weight α_m with α_m^β, where β ($0 < \beta \leq 1$) is the so-called AdaBoost-beta. A small value for β reduces the risk of overtraining because it takes longer to train on statistical fluctuations but may reduce the performance
AdaBoost performs best on weak classifiers. This means that the maximum depth of decision trees should be limited to low numbers, if they are used in combination with AdaBoost. Fig. 8.2 shows the test error, a quantity equivalent to the misclassification error but evaluated on an independent test set, as a function of the number of boosting iterations. The decision tree which is used has a depth of only one. Its error rate is around 0.45, which is only slightly better than random guessing, where the error rate is 0.5. When boosting is used, the error rate decreases to only 0.05, which is a huge improvement. The boosted decision tree also performs better than one single decision tree with 244 nodes, whose error rate is still at 0.25. After approximately 300 boosting iterations there is only little or no performance gain at all.

8.3.2. A General Approach to Boosting

The concept of boosting can be generalized. Building a committee of decision trees like in Eq. (8.9) can be interpreted as fitting an additive expansion in a set of

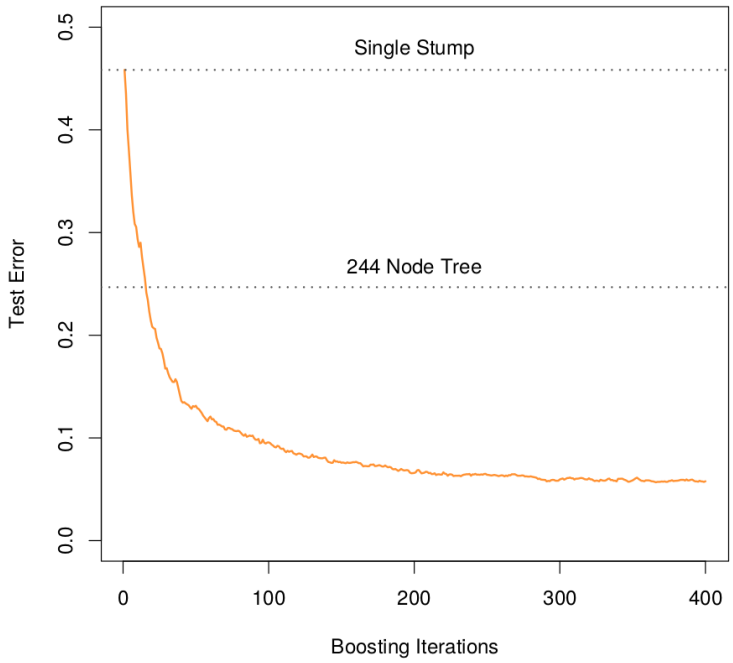


Figure 8.2.: Test error of a decision tree with a depth of one, boosted with AdaBoost, as the function of the number of boosting iterations. The test errors of the single decision tree (Single Stump) and a decision tree with 244 nodes (244 Node Tree) are also indicated [128].

functions which build a basis. For AdaBoost those basis functions are the individual decision trees $G_m(x)$. A general basis function expansion can be written as

$$f(x) = \sum_{m=1}^{M} \beta_m b(x; \gamma_m) \,, \tag{8.10}$$

where $f(x)$ is the function to approximate, β_m the expansion coefficients, and $b(x; \gamma_m)$ the basis function with the input observables x as argument and γ_m as parameter.
To obtain the best approximation the optimal values of β_m and γ_m need to be found. A measure to quantify the agreement between the model and the training data is needed. Functions which provide such a measure are called *loss functions* and are denoted here as $L(y, f(x))$. They take the output values, y, and the approximated function, $f(x)$, as arguments. Loss functions are discussed below in more detail.
The optimal values for β_m and γ_m can be found by minimizing the loss function:

$$\min_{\{\beta_m, \gamma_m\}_{m=1}^{M}} \sum_{i=1}^{N} L\left(y_i, \sum_{m=1}^{M} \beta_m b(x_i; \gamma_m)\right) \,. \tag{8.11}$$

Forward Stagewise Additive Modeling

The optimization described in Eq. (8.11) is computationally very expensive, since this is a optimization in a high-dimensional space. However, the optimization can

be approximated with an iterative approach, known as *forward stagewise additive modeling*. If the approximation up to step $m-1$ is already known, the parameters and coefficient of step m can be found in the following way, while keeping the parameters and coefficients of the previous basis functions constant:

$$(\beta_m, \gamma_m) = \arg\min_{\beta,\gamma} \sum_{i=1}^{N} L(y_i, f_{m-1}(x_i) + \beta b(x_i, \gamma)) \,. \tag{8.12}$$

After finding β_m and γ_m the approximated function at step m can be build with

$$f_m(x) = f_{m-1}(x) + \beta_m b(x_i, \gamma_m) \tag{8.13}$$

The first basis function, f_0, is usually initialized with zero, i.e. $f_0(x) = 0$.
It can be shown that the AdaBoost algorithm is equivalent to forward stagewise additive modeling if an exponential loss function,

$$L(y, f(x)) = \exp\left[-yf(x)\right] \,, \tag{8.14}$$

is used [128]. The basis functions are here the individual decision trees, $G_m(x)$. This connection is not trivial and was only observed several years after the AdaBoost algorithm was developed.

Loss functions

Loss functions quantify the discrepancy between a model and the correct output value. They depend on the margin, which is defined as $y \cdot f(x)$. The margin in classification problems has the analogous role of the residual $y - f(x)$ in regressions. A positive margin indicates that the event was classified correctly, a negative margin signals misclassification. The boundary between an correctly classified event and misclassified event is at $f(x) = 0$.
Loss functions are most of the time monotone decreasing functions, so that misclassified events are penalized by a high value of the loss function. A list of common loss functions is given below.

- Misclassification: $I(\text{sign}(f) \neq y)$
- Exponential: $\exp(-yf)$
- Binomial Deviance: $\log(1 + \exp(-2yf))$
- Squared Error: $(y - f)^2$
- Support Vector[2]: $(1 - yf)_+$

They are shown as a function of the margin in Fig. 8.3.
Both the exponential loss and binomial deviance can be seen as a continuous approximation of the misclassification loss, which is a step function. For an

[2]The label $_+$ denotes that negative values are set to zero.

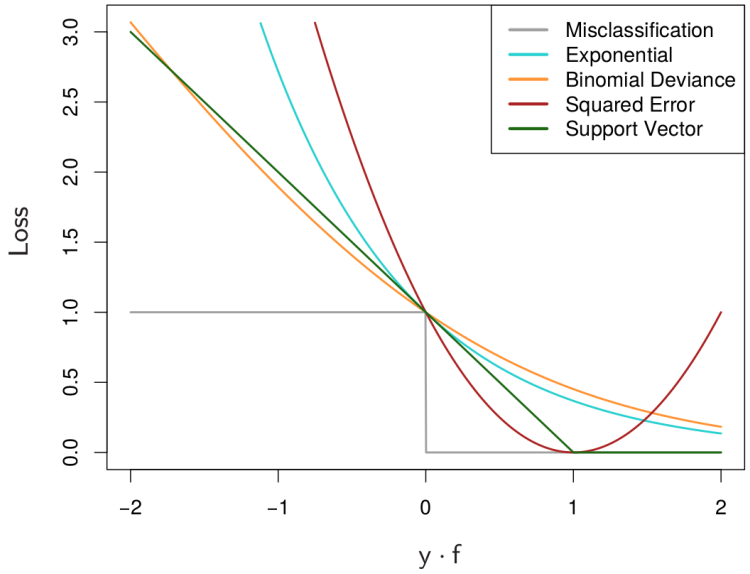


Figure 8.3.: Value of different loss functions as a function of the margin $y \cdot f$. The functions are scaled in such a way that they pass through $(0, 1)$ [128].

increasing negative margin the penalty of the binomial deviance grows in a linear fashion, for the exponential loss it grows exponentially. The binomial-deviance loss is more robust compared to the exponential loss, because misclassified events have less impact. The squared-error loss is not a good replacement for the misclassification loss, since correctly classified events above a margin of one receive an increased penalty.

8.3.3. Gradient Boost

The *gradient boost* algorithm is a boosting algorithm which can use any differentiable loss function. It applies the approach of forward stagewise additive modeling. The loss function is minimized via the *gradient descent* algorithm, a popular algorithm used to find the minimum of a function, in function space. The gradient, which is used for the minimization, is calculated as follows:

$$g_{im} = \left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}\right]_{f(x_i)=f_{m-1}(x_i)}. \tag{8.15}$$

However, this leads to an problem. The gradient needs to be known for every point in the space of the input observables, x_i. But since those are provided from a finite amount of training data, the gradient is only known at those points. The solution is to use a regression tree for the estimation of values between the training data points.
The learning rate can be adjusted in a similar way like for the AdaBoost algorithm. A parameter ν $(0 < \nu \leq 1)$ called *shrinkage* is introduced, which acts as the

learning rate. The output of every decision tree is scaled with the shrinkage,

$$f_m(x) = f_{m-1}(x) + \nu\beta_m b(x_i;\gamma_m)\,. \tag{8.16}$$

8.4. Hyperparameters

Throughout this chapter several parameters of boosted decision trees are mentioned. Those parameters are not determined automatically and need to be set by the user. In the machine-learning community, they are also known as *hyperparameters*. A large part of the next chapter is dedicated to find optimal values for those parameters. This section gives an overview of the hyperparameters and a general recommendation which values lead to the best performance. For some parameters a short identification string is introduced, which makes it easier to refer to those parameters later on.
To grow a single decision tree different thresholds for the observables are tested. The number of grid points in the range of the observable needs to be set. Studies suggest that a value of 20 is high enough to get a good splitting performance. A notable increase of the separation power with a higher number of grid points could not be seen [127]. Therefore, this analysis does not optimize the number of grid points but uses the suggested value.
For the cancellation criteria the maximum depth of the tree (`MaxDepth`) and the minimum number of events in one node given as the percentage w.r.t. the total event count (`MinNodeSize`) are used. Since boosting works best with weak classifiers, a small amount of trees should be chosen. To prevent overtraining the minimum number of events in one node should not be to small.
For boosting first it has to be decided which boosting algorithm (`BoostType`) should be used. In case of the gradient-boost algorithm the loss function has to be set. TMVA uses the binomial-deviance loss. Also the number of single decision trees which are used during the boosting (`NTrees`) and the learning rate (`Shrinkage` for gradient boost and `AdaBoostBeta` for AdaBoost) need to be set. Generally a large number of trees and a low learning rate are preferred, since they decrease the chance of overtraining and lead to a lower misclassification rate.

9. Multivariate Analysis

In this chapter a method for applying boosted decision trees to the $H \to \tau_{\text{lep}}\tau_{\text{lep}}$ analysis is presented. First, the event selection of the cut-based analysis (CBA) is modified to allow for more inclusive regions. Afterwards the BDT training and evaluation scheme is introduced in Sections 9.2 and 9.3, which is then applied to the optimization of the hyperparameters and set of observables used for the BDTs (Section 9.4). The chapter closes with evaluation the modeling of the chosen observables and BDT output distributions.

9.1. Event Selection

The boosted decision trees could be used to separate signal and background events in the VBF and boosted category as defined in Section 6.4, to increase the sensitivity. However, a lot of signal and background events are already discarded by all the thresholds set on observables in the preselection and categorization step. These thresholds were used to select the signal topology and to suppress background contributions. But BDTs offer an alternative way to separate between signal and background events, which is more powerful because no event is discarded. Therefore, some thresholds which are used to suppress background are removed to increase the signal yield. The increased number of background events is handled by the BDTs, so that in the end the sensitivity should increase. Nevertheless, removing too many thresholds will result in the opposite effect, because the training of decision trees focuses first on background processes, which are easy to separate from the signal. Because the depth of the decision trees is limited, not the full splitting potential for background events which are harder to separate from signal events may be reached. Thus, requirements on observables which remove huge amounts of background while increasing the signal-to-background ratio are not modified.
Only thresholds after the requirement of opposite charge of the two decay leptons (Cut 6) are considered for removal. The requirements on the dilepton mass, missing transverse energy, and missing transverse energy for high-p_{T} objects (Cut 7, Cut 9, Cut 10) are kept, because they suppress a large amount of $Z/\gamma^* \to \ell\ell$ background without removing a lot of signal events. Furthermore, the threshold on the transverse momentum of the leading jet (Cut 8) is also kept, since it helps to select both the VBF and boosted topology. The thresholds on the momentum fractions $x_{1,2}$ and the angular differences $\Delta\eta_{\ell\ell}$ and $\Delta R_{\ell\ell}$ (Cut 11, Cut 12, Cut 13) are removed, since they remove a sizable amount of signal events without increasing the signal-to-background ratio. The requirement on the collinear mass (Cut 14) has to be kept, otherwise the orthogonality to the $H \to WW$ analysis would be

violated. Since the output of the missing mass calculator, m_{MMC}, will be used as an input variable for the BDTs, only events can be used where the MMC mass reconstruction algorithm did converge. Thus, Cut 15 is also kept. Events containing b-jets are still rejected (Cut 16), since this provides a good discrimination against the single-top and $t\bar{t}$ background.
In the definition of the VBF category only the requirement that the two jets need to occupy different hemispheres (Cut 2V) is dropped, because for a strongly boosted system the jets can also be in the same hemisphere. The boosted category is not directly modified, but affected by the change in the VBF category, because a veto on events which pass the VBF selection is used.
Because the thresholds are changed, the pre-fit normalization factors which are discussed in Section 7.2 need to be recalculated. The normalization factors are calculated both in the modified VBF and boosted category. The results are shown in Table 9.1. In the following these normalization factors for the individual categories are used.

Table 9.1.: Pre-fit normalization factors for the multivariate analysis.

Selection	$Z \to \ell\ell$	$Z \to \tau\tau$	Top
VBF category	0.97 ± 0.15	1.07 ± 0.15	1.14 ± 0.03
Boosted category	1.04 ± 0.03	1.05 ± 0.03	1.08 ± 0.01

Furthermore, the categories are split into subcategories based on the flavour combination of the final state leptons. The SF category holds events where the final state leptons have the same flavour (ee and $\mu\mu$). Events with $e\mu$ and μe are sorted into the DF (*different flavour*) category. This increases the sensitivity and leads to a better training of the BDT, since those regions have a different background composition and signal-to-background ratio. The event yields in each subcategory are given in Table 9.2.
For training BDTs it is also important that the number of simulated events is not too low. The number of simulated events for each subcategory is shown in Table 9.3. The training statistics for the ggF and VBF signal process are very good. Other signal processes are not used in the training, as discussed in Section 9.4. Also most background processes have at least 1000 simulated events left.
There are around 50 % more signal events in the VBF and boosted category of the multivariate analysis than for the cut-based analysis. However, the background yields also increase by a factor of 3 and 1.5 in the VBF and boosted category, respectively. This reduced the signal-to-background ratio of 3.7 % and 1.1 % the VBF and boosted category of the cut-based analysis to 2.2 % and 0.97 %, respectively.

Table 9.2.: Expected event yields for the different signal and background processes in the subcategories of the multivariate analysis with a combined 2016 and 2016 dataset of 36.1 fb^{-1}. Normalization factors are applied on the top-quark, $Z \to \ell\ell$, and $Z \to \tau\tau$ background. Only statistical uncertainties are shown.

Process	VBF category		Boosted category	
	SF	DF	SF	DF
ggF $H \to \tau\tau$	3.61 ± 0.13	8.85 ± 0.21	33.22 ± 0.36	53.86 ± 0.47
VBF $H \to \tau\tau$	9.04 ± 0.08	17.01 ± 0.11	7.16 ± 0.07	10.69 ± 0.09
WH $H \to \tau\tau$	0.09 ± 0.02	0.17 ± 0.03	2.82 ± 0.14	4.43 ± 0.18
ZH $H \to \tau\tau$	0.02 ± 0.01	0.10 ± 0.02	1.67 ± 0.08	2.11 ± 0.10
ttH $H \to \tau\tau$	0.07 ± 0.04	0.32 ± 0.10	1.94 ± 0.18	3.52 ± 0.24
Fakes	73 ± 13	293 ± 16	713 ± 36	987 ± 31
Top	74 ± 4	251 ± 7	511 ± 10	1368 ± 16
Diboson	22.30 ± 0.91	64.58 ± 1.92	306.60 ± 4.22	671 ± 14
$Z/\gamma^* \to \ell\ell$	211 ± 90	10.52 ± 3.72	1082 ± 101	88 ± 11
$Z/\gamma^* \to \tau\tau$	187 ± 9	574 ± 17	2608 ± 37	4365.83 ± 42
$H \to WW$	7.47 ± 0.36	13.60 ± 0.47	24.17 ± 0.91	40.79 ± 1.28
Total signal	12.83 ± 0.16	26.46 ± 0.26	46.80 ± 0.44	74.61 ± 0.58
Total background	575 ± 92	1207 ± 25	5244 ± 114	7521 ± 58

Table 9.3.: Number of simulated events for the different signal and background processes in the subcategories of the multivariate analysis.

Process	VBF category		Boosted category	
	SF	DF	SF	DF
ggF $H \to \tau\tau$	1137	2614	11 111	17 433
VBF $H \to \tau\tau$	16 613	31 003	13 304	19 654
WH $H \to \tau\tau$	18	34	550	835
ZH $H \to \tau\tau$	6	29	480	611
ttH $H \to \tau\tau$	36	94	659	1367
Fakes	786	3130	6602	17 807
Top	525	1739	3918	10 324
Diboson	2633	7007	32 222	66 007
$Z/\gamma^* \to \ell\ell$	1447	168	11 463	1233
$Z/\gamma^* \to \tau\tau$	3533	9142	57 313	90 378
$H \to WW$	1026	1775	1418	2210
Total signal	17 810	33 770	26 104	39 900
Total background	9950	22 961	112 936	187 959

9.2. Model Selection and Assessment

Boosted decision trees need some data where the correct classification is known for the training. This knowledge can only be provided by simulated events. Ideally, all simulated events are used for the training, because a higher number of training events leads to a better performing model. However, the BDT hyperparameters and input variables need to be optimized. A way of estimating the performance of a specific BDT is needed.
The performance of a BDT cannot be measured with the same set of simulated events which was also used to train this BDT, because this would introduce a bias. An independent set of simulated events, the so-called *validation* set, needs to be used to estimate the performance of the BDT.
In high-energy physics analyses simulated events are needed for background estimation. For this neither the events from the training set and validation set can be used, because this would again introduce a bias. A third set of simulated events, the *test set*, is needed for the background estimation, which has to be independent from the two other sets.
The amount of simulated events is not unlimited, and especially the uncertainty of the measurement (where the test set is used) depends on the number of simulated events. On the other side the training statistics should also not be chosen to small, so that the performance of the BDT does not suffer. Because the training, validation, and test set need to be independent of each other there is always a tradeoff between the contributions to those sets. In a typical splitting scheme 50 % of the events are used for the training set, 25 % for the validation set, and 25 % for the test set [128].

9.3. k-fold Cross-Validation

The k-fold cross-validation approach [128] is one solution to improve the statistics for the training, validation, and test set. Here the full set of simulated events is split into k slices of equal size. Usual values for k are 5 or 10 [128]. In this analysis $k = 10$ is used. There should be no or only a very small dependence on the choice of k. The splitting is done with the help of a random number, which is generated once for each event.
Now k different BDTs are trained, each using $k - 2$ slices for training, one slice for validation, and one slice for testing. If the slices are distributed correctly as illustrated in Fig. 9.1 every slice is once used for validation and testing and $k - 2$ times for training. This improves the amount of training statistics for each BDT. For $k > 4$ the fraction of events which are used for training is always bigger than 50 %.
The validation and testing step is performed on the full set of simulated events by combining all k BDTs. For each event there are exactly 2 BDTs which were not trained with this event. One BDT is used for validation and the other one for testing.

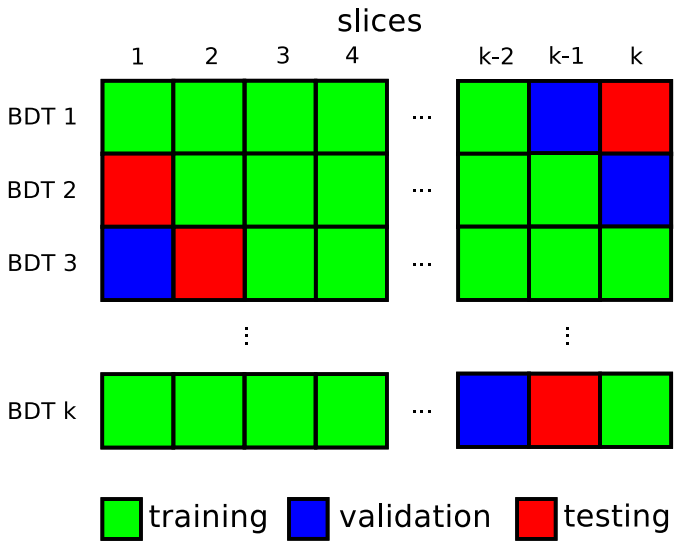


Figure 9.1.: An illustration of the k-fold cross-validation scheme. The total set of simulated events is split into k slices. All k BDTs are trained, validated, and tested by using different combinations of the slices.

In the testing stage the k BDTs need also be applied to data events. The same approach as for simulated events could be used, where a random number decides which BDT is used for which slice (random splitting). However, some scientists do not like the idea that data are treated randomly. Furthermore, random numbers cannot be reproduced on all computer systems, even if they are seeded. In this case for each data event the output of every BDT could be calculated and an average over all BDT outputs could be built (denoted as *average*). But this leads to another issue. First, the simulated events and data events are treated in a different way. Second, averaging over k BDTs could lead to the effect, that events in border regions are shifted towards the middle of the distribution, since the central limit theorem can be applied here.
In this analysis a third approach is used, the so-called *modulo splitting*. Every recorded data event in ATLAS is labeled with a unique number, the so-called event number. This number is set once and not modified again. The expression "event number mod k" is used to split the data events into k different slices. This method was chosen because it treats data events similar to simulated events, but takes out the randomness of the splitting. It needs to be checked that the data events are distributed equally in the different slices. The distribution of "event number mod k" is shown in Fig. 9.2 for all data events which pass the MVA selection. Within two standard deviations the count in each slice agrees with the average. Therefore this splitting procedure does not introduce slices of unequal size.
The three methods of data treatment are compared in Fig. 9.3. Here the final BDTs which are selected in Section 9.4 are used. In the low-BDT-score regions in the boosted category the shift of events in the border region towards the center when averaging over all BDTs can be seen. Otherwise the methods agree within

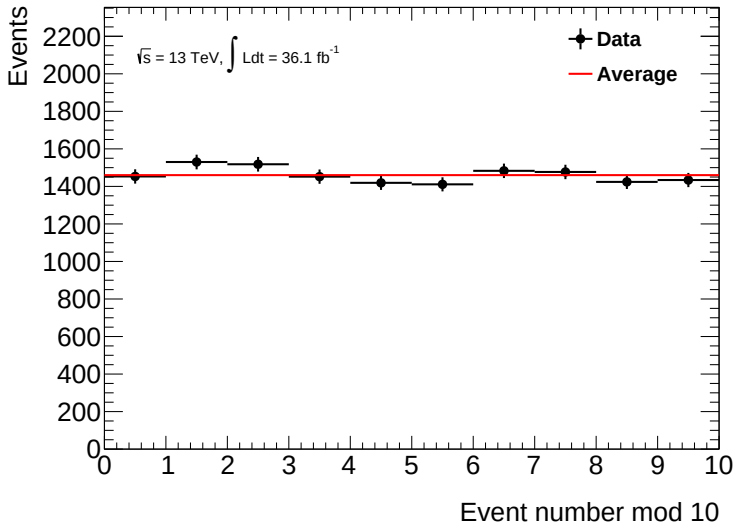


Figure 9.2.: Number of data events for each slice when splitting with "event number mod k" is used. The plot shows data from the full 2015 and 2016 dataset which passed the MVA selection.

uncertainties.

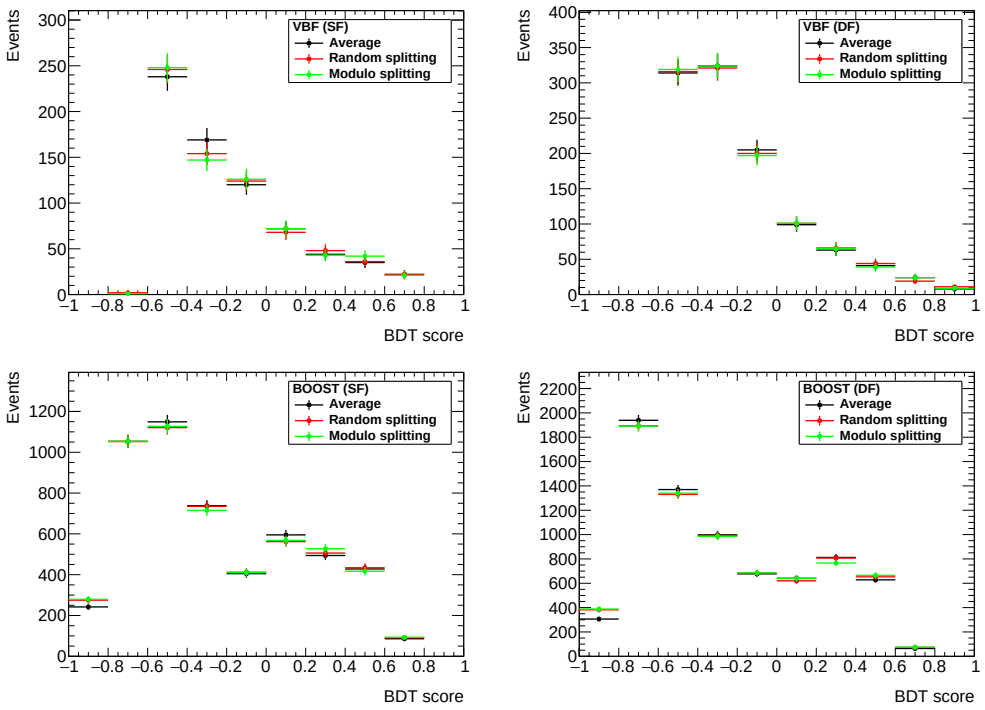


Figure 9.3.: Comparison of data treatment in k-fold cross-validation in the four MVA categories. The optimized BDTs from Section 9.4 are used and are evaluated on the full 2015 and 2016 dataset.

9.4. Optimization

The optimization is done individually in the four signal regions (VBF SF, VBF DF, boosted SF, boosted DF), which are defined in Section 9.1. In the VBF category only the VBF $H \to \tau\tau$ sample is used for training and the boosted category uses only the ggF $H \to \tau\tau$ sample as signal. The other signal processes are discarded. This decision was made even though the contributions of the VBF process in the boosted category and the ggF process in the VBF category are not negligible. Since the goal of this thesis is to measure the signal strength of $H \to \tau\tau$ in the VBF- and ggF-production channel, it makes sense to optimize the analysis for the measurement of the individual production modes.

The optimization procedure is divided into two steps. First, the BDT hyperparameters are optimized. Here 54 input variables are used. A full list of the considered observables can be found in Appendix A.1. However, such a high number of observables is not preferred, since the modelling of every observable needs to be tested. Some observables are also highly correlated and provide only a tiny amount of new information. Thus, the number of input variables for the BDTs is reduced in a second step, keeping only the variables which provide the highest separation power.

9.4.1. Figure of merit

The separation power of a BDT can be assessed in different ways. In this section several possible figures of merit are discussed, which were considered for the estimation of the BDT performance. These values are calculated on the validation set.
A common characteristic of a machine-learning model is the area under the receiver-operating characteristic (ROC) curve. The ROC curve displays the rate of background rejection as a function of the rate of signal efficiency. A larger area under the ROC curve indicates a better separation power. The ROC curves are provided by TMVA.
Another figure of merit is the separation $\left\langle S^2 \right\rangle$, which is defined by [127]

$$\left\langle S^2 \right\rangle = \int_{-1}^{1} \frac{\left(\hat{y}_S(y) - \hat{y}_B(y)\right)^2}{\hat{y}_S(y) + \hat{y}_B(y)} \, \mathrm{d}y \,. \tag{9.1}$$

The probability density functions of the output of the classifier are denoted as $\hat{y}_S$ and $\hat{y}_B$. If the signal and background distribution have a complete overlap, the separation is zero. Distributions with no overlap at all yield a separation of one. The separation is also calculated by TMVA.
TMVA also provides a significance, which is calculated by

$$Z_{\mathrm{TMVA}} = \frac{\overline{y}_S - \overline{y}_B}{\mathrm{RMS}_S(y)^2 + \mathrm{RMS}_B(y)^2} \tag{9.2}$$

Here $\overline{y}_S$ and $\overline{y}_B$ are the means of the classifier output for signal and background, respectively. The root-mean-squares of the classifier output for signal and background are denoted as $\mathrm{RMS}_S(y)$ and RMS_B.
Another way to calculate a significance is with the *binned significance*. For this histograms with 10 equidistant bins of the BDT distribution for signal and background is used. If a bin contains less than 10 background events, it is merged with its left neighbor (the most left bin is merged with its right neighbor). In each bin the significance is calculated with the asymptotic formula [130]

$$Z_{\mathrm{asym}}(s,b) = \sqrt{2\left((s+b)\ln\left(1+\frac{s}{b}\right) - s\right)}\,, \tag{9.3}$$

where s and b are the expected signal and background yields. The binned significance Z_{binned} is the quadratic sum of the significances in the individual bins,

$$Z_{\mathrm{binned}} = \sqrt{\sum_i Z_{\mathrm{asym}}(s_i, b_i)^2}\,. \tag{9.4}$$

This significance is a simple approximation of the significance of the complete fit model, which is discussed in Chapter 11.
Of course the significance of the fit model can also be used to estimate the performance of a BDT. This figure-of-merit is actually used for the optimization,

since it the closest approximation of the real fit. This method is actually used for the optimization. However, not the full fit model is applied. Including all systematic variations (Chapter 10) would need too much computing power to run the optimization in a reasonable timescale. Therefore, the fit is done without the systematic variations, which is also denoted as *stat. only fit* ("statistics only"). The fit is performed with Asimov data, i.e. the simulated events are used instead of the measured data, to avoid biasing the optimization due to the observed data.

9.4.2. Hyperparameters

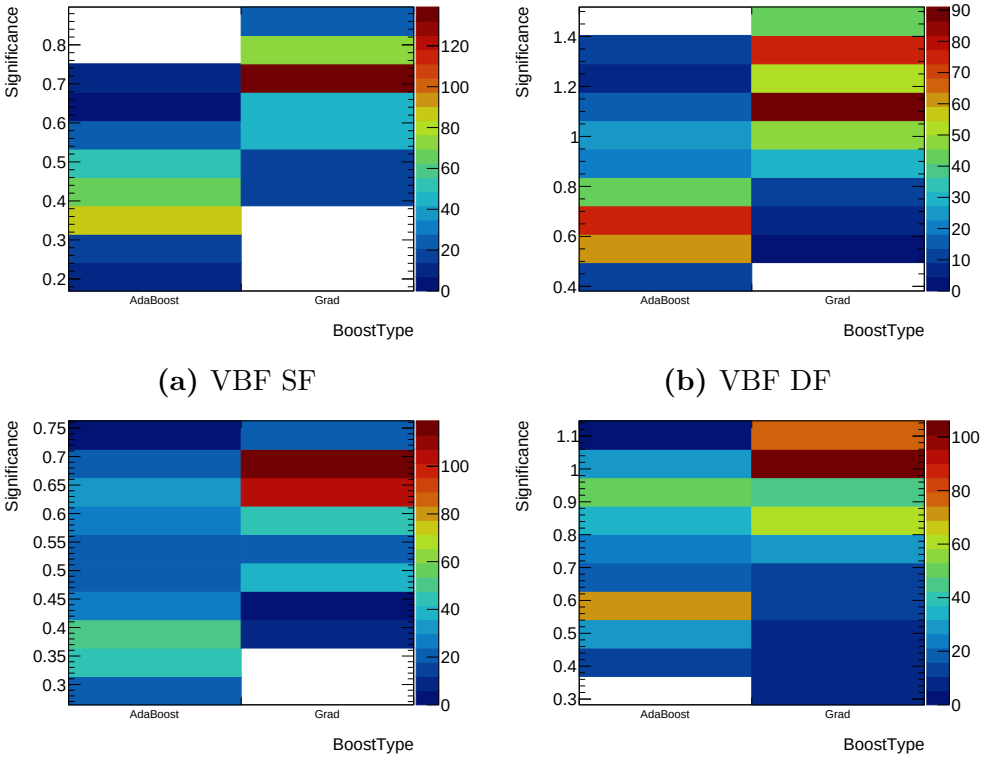
In the first optimization step different hyperparameters for the BDTs are optimized. The boosting algorithm, number of trees in the boosting, maximum depth of the tree, minimum number of events in the final nodes, and the learning rate are varied in a grid scan. The values which were considered for those parameters can be found in Table 9.4. This leads to a total of 630 BDTs which need to be trained for each region.

Table 9.4.: Values of the BDT hyperparameters which are used in the first optimization step. The hyperparameters are explained in Section 8.4.

Hyperparameter	Values
BoostType	AdaBoost, Grad (gradient boost)
NTrees	$50, 250, 500, 750, 1000$
MaxDepth	$2, 3, 4, 5, 7, 10$
MinNodeSize	$1\,\%, 5\,\%, 10\,\%$
Shrinkage	$0.05, 0.1, 0.2, 0.5$
AdaBoostBeta	$0.1, 0.5, 0.8$

General observations

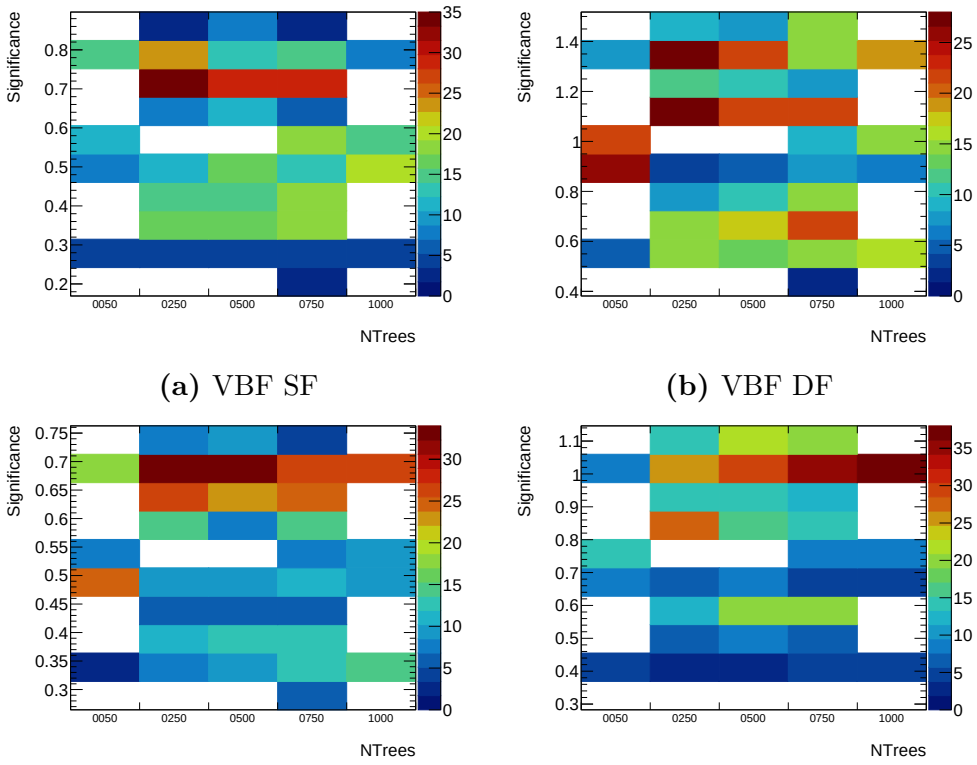
Before the best BDT hyperparameters are chosen first a few general trends are discussed. The significance of the stat. only Asimov fit depending on the boosting algorithm is shown in Fig. 9.4. In all four regions the gradient-boosting algorithm yields on average a better significance than the AdaBoost algorithm. For the VBF categories between 250 and 750 boosting iterations are preferred, in the boosted categories also BDTs with 1000 boosting iterations yield a comparable significance, as can be seen in Fig. 9.5. In the VBF SF and boosted DF category a maximum depth of at least 4 yields the best significance, as shown in Fig. 9.6. In the boosted DF category also BDTs with a maximum depth of 3 return a good significance. A general dependence of the separation power of the BDTs on the minimum number of events in the final nodes could not be observed (Fig. 9.7). A lower training rate results most of the time in a higher significance, as can be seen in Figs. 9.8 and 9.9.



(a) VBF SF **(b)** VBF DF

(c) Boosted SF **(d)** Boosted DF

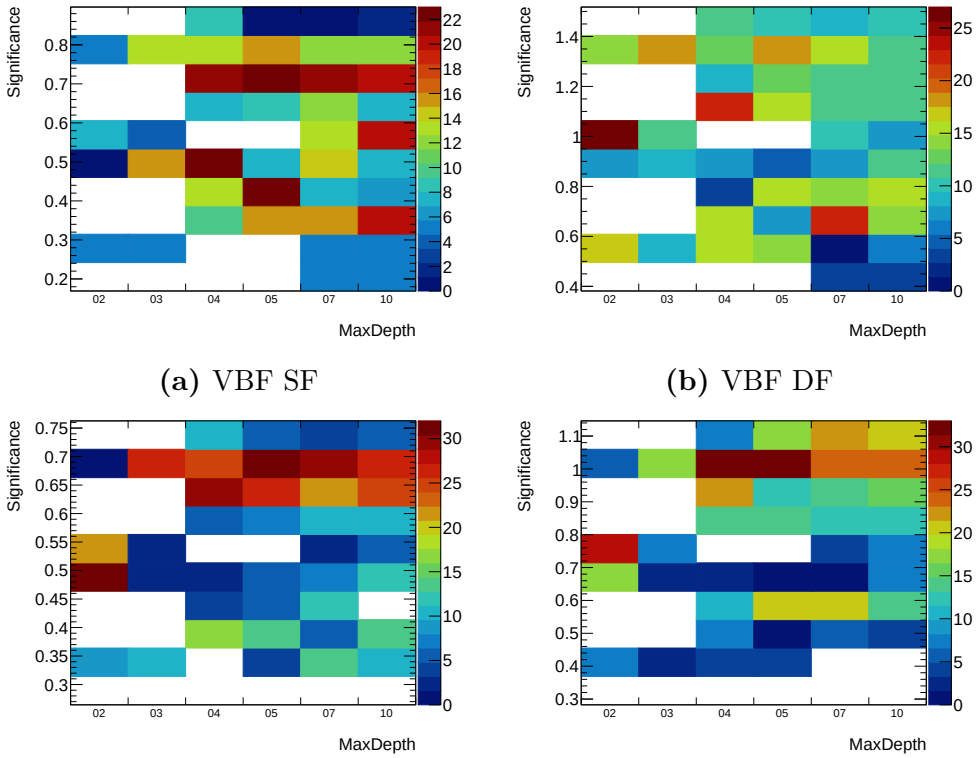
Figure 9.4.: Significance of all trained BDTs depending on the boosting algorithm for each region.



(a) VBF SF **(b)** VBF DF

(c) Boosted SF **(d)** Boosted DF

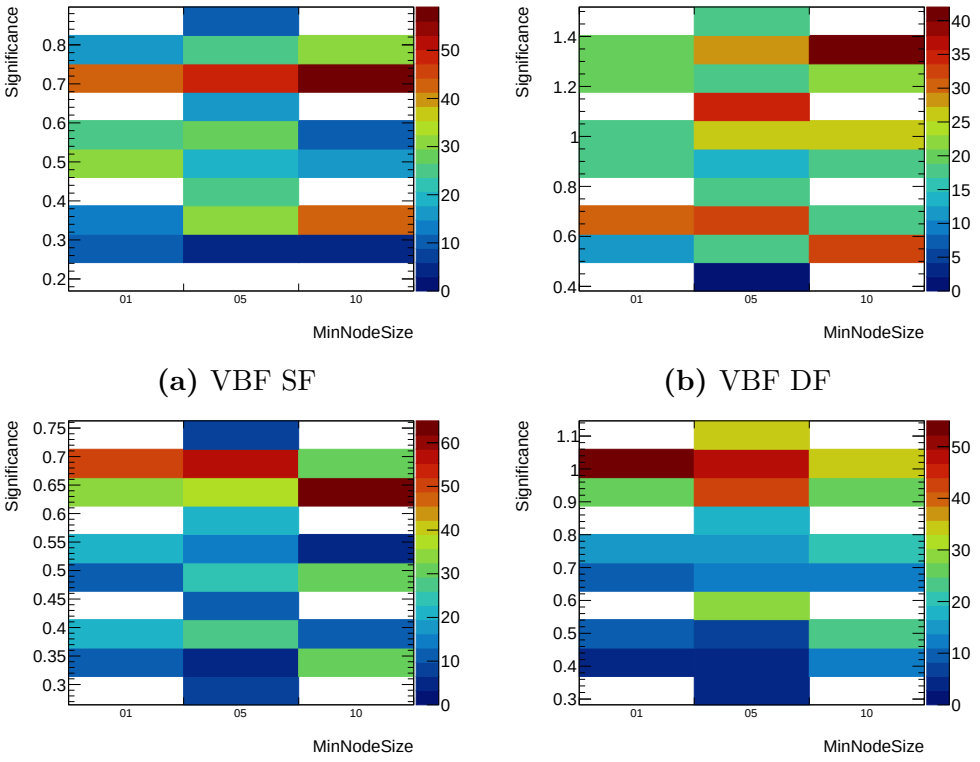
Figure 9.5.: Significance of all trained BDTs depending on the number of trees used in boosting for each region.



(a) VBF SF **(b)** VBF DF

(c) Boosted SF **(d)** Boosted DF

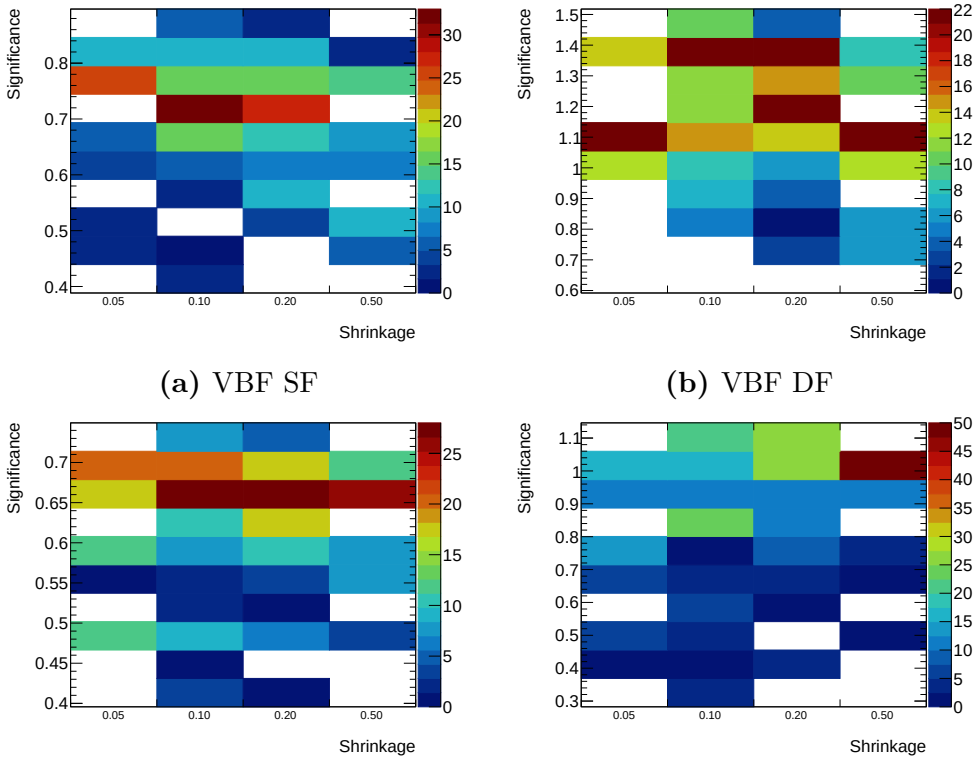
Figure 9.6.: Significance of all trained BDTs depending on the maximum depth of the decision trees for each region.



(a) VBF SF **(b)** VBF DF

(c) Boosted SF **(d)** Boosted DF

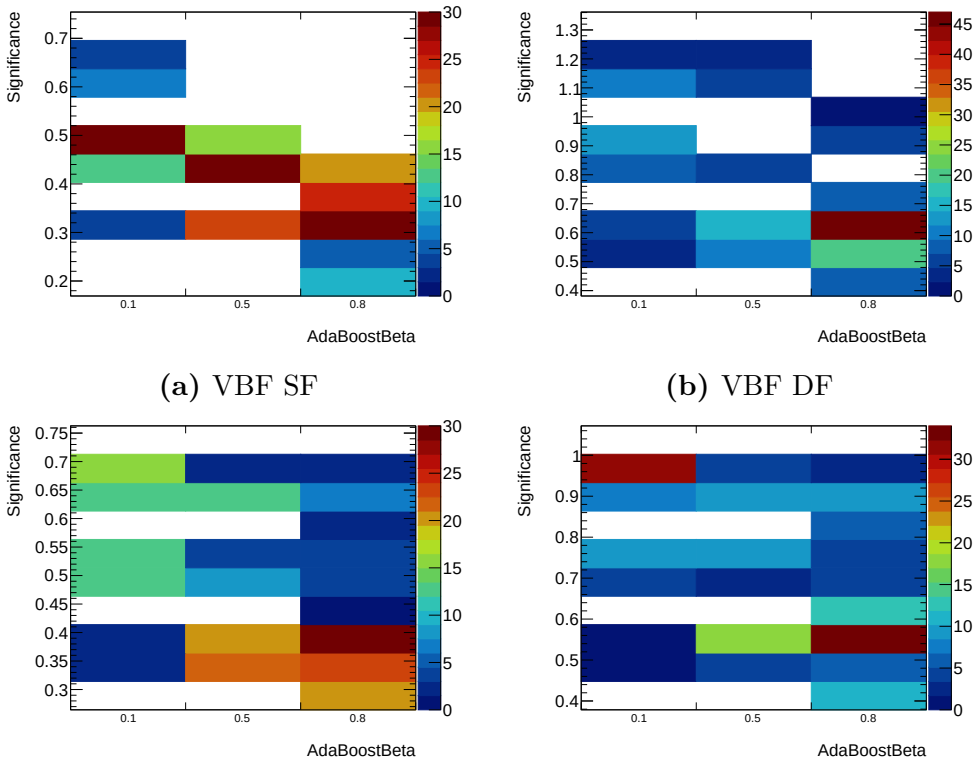
Figure 9.7.: Significance of all trained BDTs depending on the minimum number events given as the fraction of all events for each region.



(a) VBF SF **(b)** VBF DF

(c) Boosted SF **(d)** Boosted DF

Figure 9.8.: Significance of all trained BDTs where the gradient boost algorithm was used depending on the learning rate for each region.



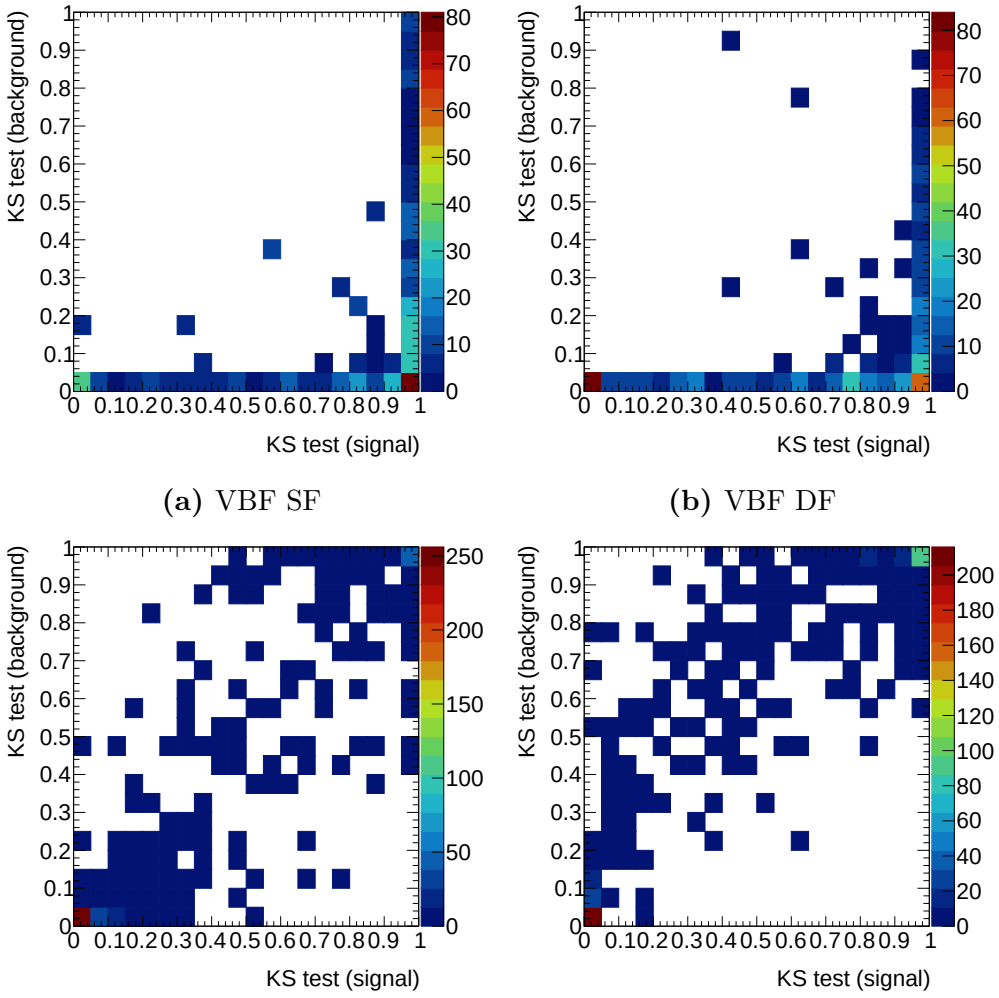
(a) VBF SF **(b)** VBF DF

(c) Boosted SF **(d)** Boosted DF

Figure 9.9.: Significance of all trained BDTs where the AdaBoost algorithm was used depending on the learning rate for each region.

Result of optimization

The overtraining of a BDT can be estimated by using the *Kolmogorov–Smirnov test* [131, 132] (KS-test). Here the BDT output of training and validation set is compared for both the signal and background distribution. The KS-test yields a probability between zero and one, where one indicates perfect agreement and zero no agreement at all. Only BDTs which yield a KS-test probability for both signal and background distribution above 0.4 are considered. The exact choice of this threshold plays only a very minor role, since most BDTs have a KS-test probability of either one or zero, as can be seen in Fig. 9.10.



(a) VBF SF **(b)** VBF DF

(c) Boosted SF **(d)** Boosted DF

Figure 9.10.: Correlation plots of the KS-test propability between the BDT output on the training and validaiton set for the signal and background distribution.

The BDT with the best significance is selected for each region, the hyperparameters of the best performing BDTs are listed in Table 9.5. In all regions the gradient

boost algorithm is used. The number of trees in the boosting is very similar for all regions. Furthermore, a low maximum depth and learning rate are chosen. There is quite some spread on the minimum number of events in the final nodes, no specific value is preferred.
The distributions of the BDT response for the four select BDTs are shown in Fig. 9.11. There is a very good agreement between the BDT shapes of the training and validation set.

Table 9.5.: Hyperparameters of the best performing BDTs in each region.

Region	Type	NTrees	MaxDepth	MinNodeSize	LearnRate
VBF SF	Grad	250	2	1 %	0.1
VBF DF	Grad	250	5	10 %	0.05
Boosted SF	Grad	500	4	5 %	0.05
Boosted DF	Grad	250	5	5 %	0.1

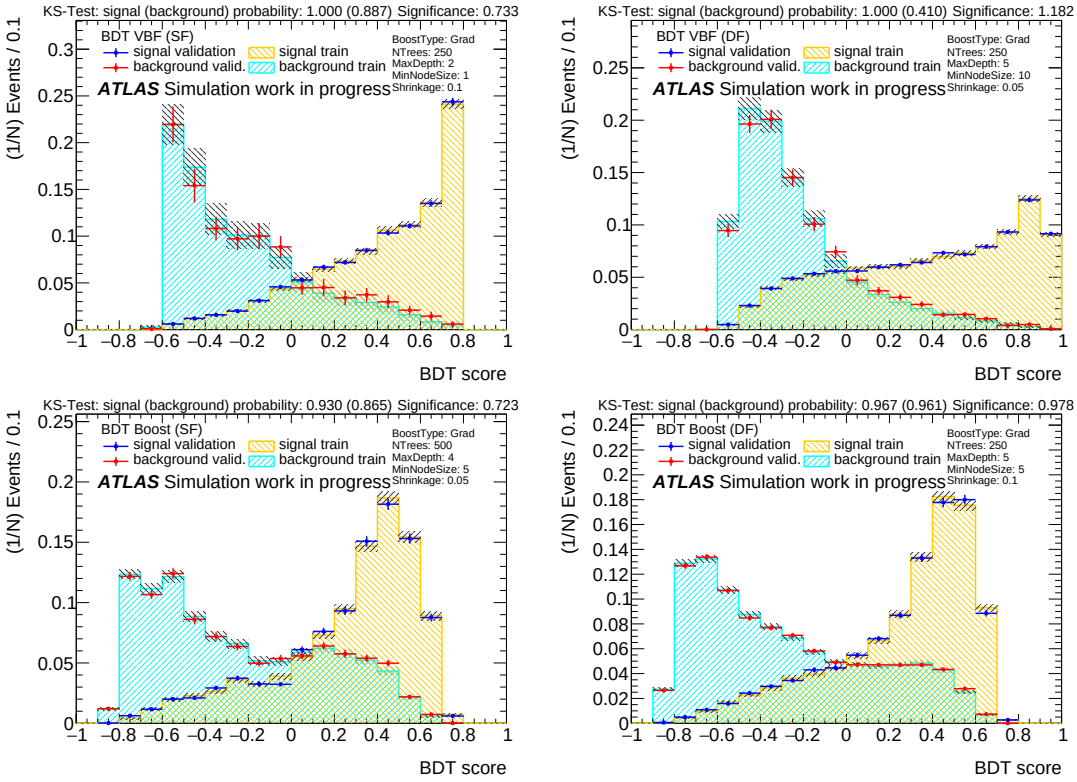


Figure 9.11.: Distributions of the best performing BDTs in the hyperparameter optimization for signal and background and the training and validation set. The top row shows the BDTs for the VBF regions and the bottom row for the boosted regions. The same flavour regions are on the left and the different flavour regions on the right. At the top of each figure the KS-test probabilities and significances can be found. The areas of all distributions are normalized to one.

9.4.3. Input variables

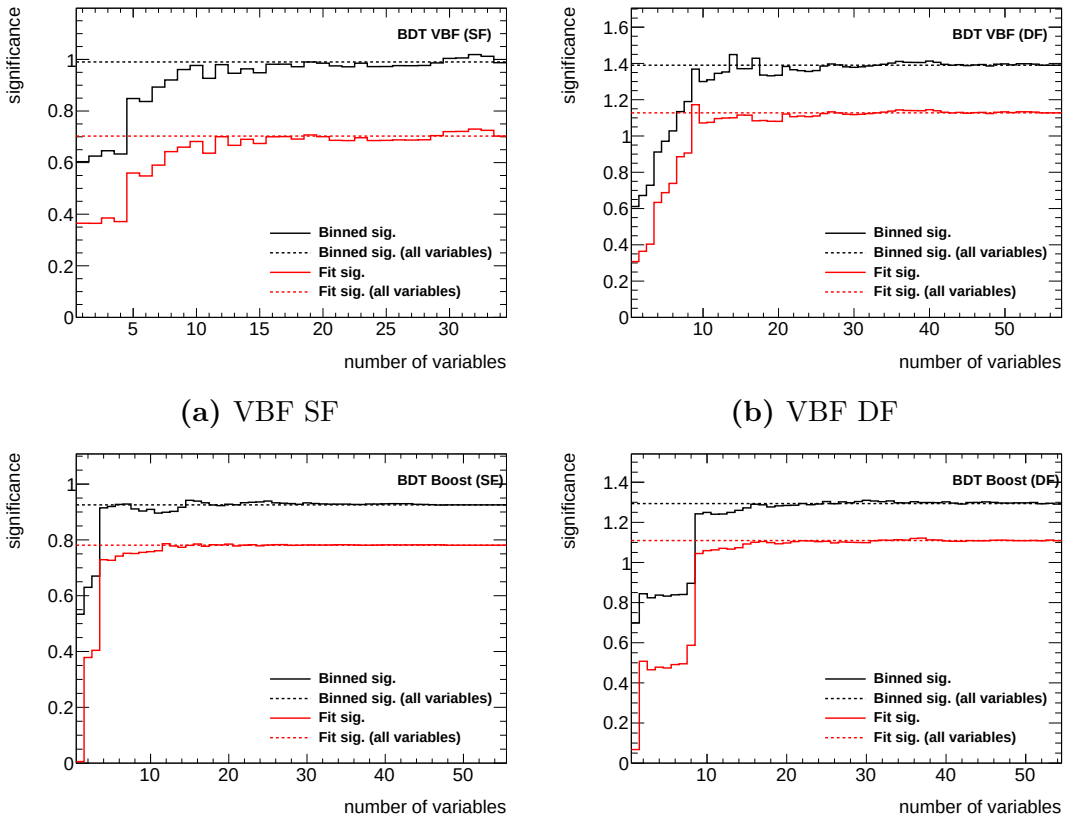
Up to now the BDTs use 54 observables as input variables. However, a low count of input variable is desired, because the modelling of every input variable has to be reviewed. Therefore, variables which have only a low impact on the separation power of the BDTs are discarded. This is done individually for each region.

The impact of an input variable on the separation power of a BDT can be estimated with the so-called *variable separation*. It is calculated by counting how often a given variable is used to split a node in the decision tree while weighting each split occurrence by the square of the gained separation ΔQ (c.f. Eq. (8.3)) and the number of events in the node [133]. This concept can also be used for the collection of decision trees in a boosted decision tree [127]. Variables which are not used at all have a value of zero. A higher value indicates that the variable is more important for the performance of the BDT. All variables used in a BDT are ranked by this number in a *variable ranking*. The variable ranking is averaged over all 10 BDTs in the k-fold cross-evaluation.

The reduction of the number of variables is based on this variable ranking. First, all variables which have a variable separation of zero are removed. Then, one by one, the least performing variable is dropped from the list of input variables. After each removal the BDT is trained again and a new variable ranking is calculated. Additionally, the significance is calculated for each BDT as described in Section 9.4.1. This iterative approach is chosen, because the variable ranking can change if one ore more variables are not used anymore.

The dependence of the significance on the number of variables is shown in Fig. 9.12 for each of the four regions. As expected, the significance does not change much at first, but after removing more and more variables it decreases. However, the curve is not always decreasing monotonically, there is some amount of statistical fluctuations caused by large event weights. This is caused by events which were generated by the Sherpa generator [80–84], as described in Section 4.3. Much more simulated events are produced than expected in collisions. To match the count of simulated and observed events, each simulated event is assigned a weight, which depends on the cross-section of the process. Due to higher order corrections it can happen that also negative events are assigned, in the case that the cross-section of the higher order correction is smaller than the leading order cross section. Single events which negative weights are unphysical, but combined with a large group of other events only the decrease of the cross-section is noticeable. There are also weights from other sources, for example from the *pile-up reweighting*. These weights can have large values. Therefore, it can happen that an event has a high negative weight. These events cause the fluctuations in Fig. 9.12. To reduce the amount of fluctuations the event weight is restricted to $-3 < \text{weight} < 1$ in the training and validation set. The upper value was also included, because it turned out that also large positive event weights cause problems.

For the BDTs in the boosted region a sharp drop-off threshold can be seen, which is used to determine the number of input variables. In the boosted SF and DF region 4 and 9 input variables is used, respectively. The decision how much variables



(a) VBF SF **(b)** VBF DF

(c) Boosted SF **(d)** Boosted DF

Figure 9.12.: Dependence of the significance on the number of input variables used in the BDT in each of the four regions.

in the VBF regions should be used was more difficult, because there is not such a clear drop-off in the expected significance. A trade-off between the decrease of significance and remaining count of variables has to be made. In the end, 9 variables were chosen for the VBF SF region and 8 for the VBF DF region. The chosen observables are discussed below. Not all observable are used in every region. The variable rankings can be found in Tables 9.6 to 9.9 Distributions of all variables are shown in the next section.

Common observables

The following variables are used in both the VBF and boosted category.

- The mass of the missing mass calculator, m_{MMC}, as discussed in Section 6.1.2.
- The missing transverse energy, $E_{\mathrm{T}}^{\mathrm{miss}}$, as defined in Section 5.6.
- The minimum $\Delta\eta$ distance between the dilepton system and all jets, min $\Delta\eta(\ell\ell, \mathrm{jets})$. This observable is not used in BDTs of the boosted SF region.

Table 9.6.: Variable ranking for the final BDT in the VBF SF region calculated by averaging over the individual variable rankings of the 10 BDTs from the k-fold cross-validation.

Rank	Observable	Mean Variable Separation	Standard Deviation
1	$\Delta R_{\ell\ell}$	0.212	0.016
2	m_{MMC}	0.178	0.015
3	$\min \Delta R(\ell\ell, \mathrm{jets})$	0.154	0.014
4	m_{jj}	0.130	0.014
5	$\min \Delta R(\ell_2, \mathrm{jets})$	0.098	0.018
6	n_{jets}	0.089	0.007
7	$E_{\mathrm{T}}^{\mathrm{miss}}\phi$ centrality	0.080	0.011
8	$p_{\mathrm{T}}^{\mathrm{total}}$	0.053	0.025
9	$E_{\mathrm{T}}^{\mathrm{miss}}$	0.005	0.007

Table 9.7.: Variable ranking for the final BDT in the VBF DF region calculated by averaging over the individual variable rankings of the 10 BDTs from the k-fold cross-validation.

Rank	Observable	Mean Variable Separation	Standard Deviation
1	$\Delta R_{\ell\ell}$	0.180	0.009
2	$\min \Delta R(\ell\ell, \mathrm{jets})$	0.165	0.009
3	$E_{\mathrm{T}}^{\mathrm{miss}}\phi$ centrality	0.144	0.009
4	m_{MMC}	0.132	0.014
5	$\min \Delta R(\ell_2, \mathrm{jets})$	0.129	0.010
6	m_{jj}	0.112	0.011
7	$p_{\mathrm{T}}^{\mathrm{total}}$	0.090	0.017
8	n_{jets}	0.048	0.009

Table 9.8.: Variable ranking for the final BDT in the boosted SF region calculated by averaging over the individual variable rankings of the 10 BDTs from the k-fold cross-validation.

Rank	Observable	Mean Variable Separation	Standard Deviation
1	m_{MMC}	0.446	0.038
2	$\Delta R_{\ell\ell}$	0.238	0.031
3	$m_{\ell\ell}$	0.227	0.019
4	$E_{\mathrm{T}}^{\mathrm{miss}}$	0.089	0.017

Table 9.9.: Variable ranking for the final BDT in the boosted DF region calculated by averaging over the individual variable rankings of the 10 BDTs from the k-fold cross-validation.

Rank	Observable	Mean Variable Separation	Standard Deviation
1	m_{MMC}	0.263	0.007
2	$m_{\text{T}}^{\ell_0}$	0.115	0.009
3	$\min \Delta R(\ell\ell, \text{jets})$	0.107	0.007
4	$m_{\ell\ell}$	0.107	0.010
5	$m_{\tau\tau,\text{j}_1}$	0.097	0.006
6	$\Delta R_{\ell\ell}$	0.085	0.009
7	$E_{\text{T}}^{\text{miss}}/p_{\text{T}}^{\ell_2}$	0.078	0.015
8	Sphericity	0.075	0.004
9	η_{ℓ_1}	0.074	0.009

- The ΔR as defined in Eq. (3.5) between the two leptons, $\Delta R_{\ell\ell}$.

VBF category specific observables

The following variables are only used in the VBF category.

- The mass of the dijet system of the two leading jets, m_{jj}.
- The number of jets. For the counting only jets with $p_{\text{T}} > 30\,\text{GeV}$ are used. Only a distinction between events with two jets and more than two jets is made, since high jet-multiplicity bins have only leading order precision.
- The $E_{\text{T}}^{\text{miss}}\phi$ centrality, which quantifies the centrality of the $\mathbf{E}_{\text{T}}^{\text{miss}}$ vector with respect to the final state leptons in the transverse plane. The transverse plane is orthogonal to the direction of both final state leptons. The smaller ϕ angle between the two leptons defines the positive quadrant. The ϕ centrality, $C_\phi(\ell\ell, k)$, is calculated for an object k by [134]

$$C_\phi^A(\ell\ell, k) = \sin(\phi_k - \phi_{\ell_1}) / \sin(\phi_{\ell_2} - \phi_{\ell_1}) \tag{9.5}$$

$$C_\phi^B(\ell\ell, k) = \sin(\phi_{\ell_2} - \phi_k) / \sin(\phi_{\ell_2} - \phi_{\ell_1}) \tag{9.6}$$

$$C_\phi(\ell\ell, k) = \frac{C_\phi^A(\ell\ell, k) + C_\phi^B(\ell\ell, k)}{\sqrt{C_\phi^A(\ell\ell, k)^2 + C_\phi^B(\ell\ell, k)^2}} \tag{9.7}$$

- The norm of the vectorial sum of the transverse momentum of the two final state leptons, the two leading jets, and the missing transverse energy, $p_{\text{T}}^{\text{total}}$.
- The minimum distance in ΔR between the leading lepton and all jets, $\min \Delta\eta(\ell_1, \text{jets})$. This observable is only used in the BDTs of the VBF DF region.

- The minimum distance in ΔR between the subleading lepton and all jets, $\min \Delta\eta(\ell_2, \text{jets})$. This observable is only used in the BDTs of the VBF SF region.

Boosted category specific observables

The following variables are only used in the boosted category.

- The mass of the dilepton system, $m_{\ell\ell}$.
- The sum of the mass of the visible decay products of the τ-lepton decay and the mass of the leading jet, $m_{\tau\tau,\mathrm{j}_1}$.
- The fraction of the missing transverse energy and the transverse momentum of the second jet, $E_{\mathrm{T}}^{\mathrm{miss}}/p_{\mathrm{T}}^{\ell_2}$. This observable is only used in the BDTs of the boosted DF region.
- The sphericity, a measure for the isotropy of the energy flow in the event [135]. First, the momentum tensor of all selected leptons and jets in the event is calculated,

$$S^{\alpha\beta} = \frac{\sum_i p_i^\alpha p_i^\beta}{\sum_i \left|\mathbf{p}_i^2\right|} \qquad \alpha, \beta = 1, 2, 3\,. \tag{9.8}$$

  The sphericity S is then built from the two smallest eigenvalues λ_2 and λ_3 of this vector,

$$S = \frac{3}{2}\left(\lambda_2 + \lambda_3\right)\,. \tag{9.9}$$

- The transverse mass of the leading lepton with the missing transverse energy, $m_{\mathrm{T}}^{\ell_1}$. This observable is only used in the BDTs of the boosted DF region.
- The pseudorapidity of the leading lepton, η_{ℓ_1}. This observable is only used in the BDTs of the boosted DF region.

Final BDT distributions

The BDT distributions of the final BDTs are shown in Fig. 9.13 for training and validation set. There is an excellent agreement between the distributions of the training and validation set for both signal and background. The correlation plots of the input variables can be found in Appendix A.2.

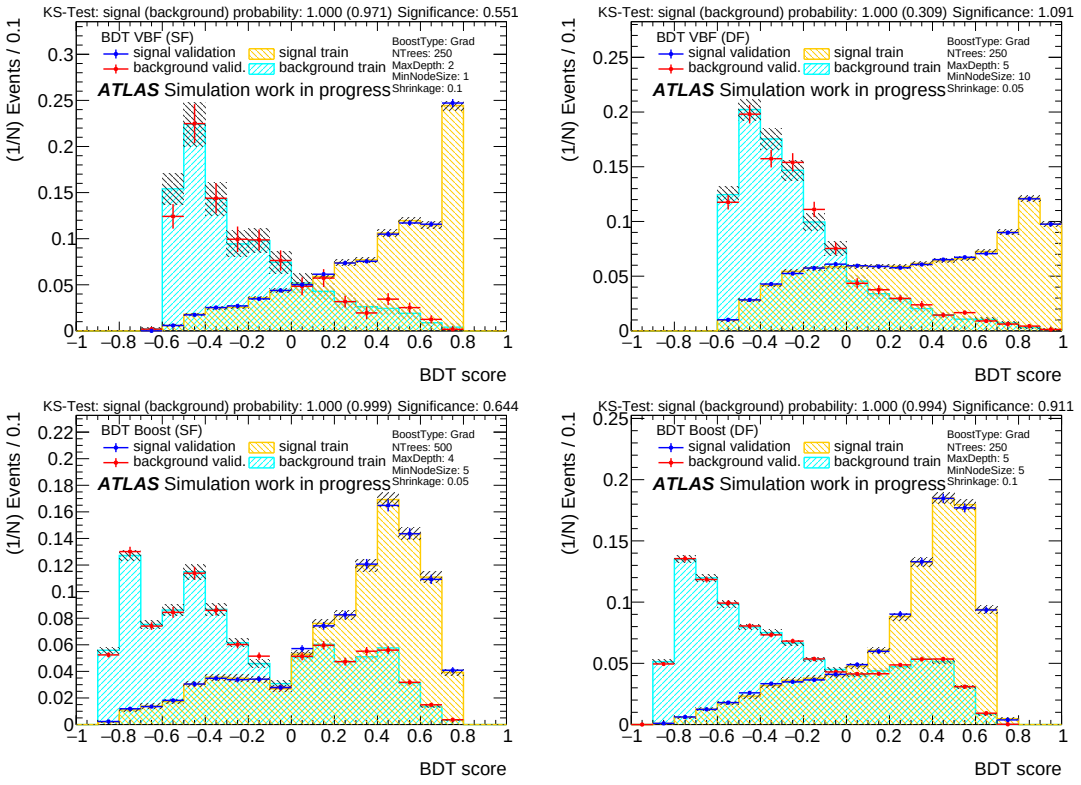


Figure 9.13.: Distributions of the final BDTs in the optimization of the number of input variables for signal and background and the training and validation set. The top row shows the BDTs for the VBF regions and the bottom row for the boosted regions. The same flavour regions are on the left and the different flavour regions on the right. The areas of all distributions are normalized to one.

9.5. Modeling of input observables

It is important that there is no mismodelling of the simulated backgrounds if compared to data events. Therefore, the modelling needs to be checked for all distributions of observables, which are used as input variables for the BDTs. Additionally the BDT output distributions themselves need also to be checked for mismodelling. Because the analysis is performed blinded, the full distributions of the BDT output and m_{MMC} can only be compared to data in the control regions, which are defined in Section 7.2. In the signal regions the m_{MMC} distributions are blinded for $100\,\mathrm{GeV} \leq m_{\mathrm{MMC}} \leq 150\,\mathrm{GeV}$, and the BDT distributions are blinded for an BDT output value greater than zero.
To quantify the agreement between the background model and observed data a χ^2-test is performed for all observables. Due to the blinding this cannot be done for the m_{MMC} distribution in the signal regions. The χ^2-test probabilities are expected to follow the uniform distribution, if the predictions of the SM are correct. Unless the χ^2-test probability is tiny, i.e. $< 10^{-3}$, the mismodelling can be accepted, since only statistical fluctuations but not systematic variations are considered.

9.5.1. Signal region

The distributions of the observables used in the BDT of the VBF SF region are shown in Fig. 9.14. Due to the events with large negative weights there is a large statistical fluctuation in the $Z \to \ell\ell$ background, which sometimes leads to bins with an expected negative event yield and large statistical errors. Some minor disagreement between expected background distributions and distributions of observed data can be seen for all observables. However, the χ^2-test probabilities, which are listed in Table 9.10, are still consistend with the exptectation of a flat distribution. For the VBF DF region, the modeling is better, since there is only a very minor contribution of the $Z \to \ell\ell$ background. This results in higher χ^2-test probabilities. Only the m_{jj} distribution shows some larger mismodelling. In the distributions of the observables used in the BDTs of the boosted categories (Figs. 9.16 and 9.17) generally a good agreement between simulated and observed events can be seen, except for the $\Delta R(\ell,\ell)$, $E_{\mathrm{T}}^{\mathrm{miss}}/p_{\mathrm{T}}^{\ell_1}$, Sphericity and $\min\Delta\eta(\ell\ell,\mathrm{jets})$ distributions. This leads to high χ^2-test probabilities, expect for the just mentioned distributions. The distributions of the BDT output in the different signal regions are shown in Fig. 9.18. For the BDTs in the VBF category a good separation between signal and background is reached. However, the background distributions of the BDTs in the boosted category start rising again for very high values of the BDT output. Additionally, there are two peaks in the distribution of the BDT for the boosted SF category. Due to the large errors in the VBF SF category the background distributions agrees with the observed data within uncertainties. A slight overshoot of simulated events can be seen in the BDT distribution of the VBF DF category. In the BDT distributions of the boosted category the data events agree with the simulated distributions within uncertainties except for one bin in the BDT of the boosted DF region.

9.5.2. Control regions

As a further validation, the distributions of observables and BDT outputs can also be checked in the control regions. The modeling in the top-quark control regions (Figs. 9.19 and 9.21) is much better than in the $Z \to \ell\ell$ control regions (Figs. 9.20 and 9.22). This can also be seen in the χ^2-test probabilities (Tables 9.12 and 9.13), where the probabilities for multiple observables are zero in the $Z \to \ell\ell$ control region. The cause are again events with high negative weights in the $Z \to \ell\ell$ background. However, in the statistical analysis only the yield in the control regions is used and not the shape, so if the overall yields are modeled well this should be no major issue. Additionally, only statistical uncertainties are considered for the χ^2-test.
The distributions of the BDT output in the different control regions are shown in Fig. 9.23. The same trend in modeling as for the input variables can be seen. In all these distributions the BDT output peaks at low values, which indicates that the BDTs indeed categorize most background events correctly.

Table 9.10.: χ^2-test probabilities between the background distributions of simulated events and data distributions for the observables used as input variables in the VBF SF and DF regions. Only statistical uncertainties are considered.

variable	χ^2 prob. SF	χ^2 prob. DF
$\min \Delta\eta(\ell\ell, \text{jets})$	0.03	0.84
m_{jj}	0.35	0.02
$\Delta R(\ell, \ell)$	0.84	0.78
n_{jets}	0.05	0.84
$E_{\text{T}}^{\text{miss}}$ Φ centrality	0.02	0.80
$p_{\text{T}}^{\text{total}}$	0.13	0.28
$E_{\text{T}}^{\text{miss}}$	0.87	–
$\min \Delta R(\ell_1, \text{jets})$	0.10	–
$\min \Delta R(\ell_0, \text{jets})$	–	0.92

Table 9.11.: χ^2-test probabilities between the background distributions of simulated events and data distributions for the observables used as input variables in the boosted SF and DF regions. Only statistical uncertainties are considered.

variable	χ^2 prob. SF	χ^2 prob. DF
$m_{\ell\ell}$	0.25	0.27
$E_{\text{T}}^{\text{miss}}$	0.61	–
$\Delta R(\ell, \ell)$	0.15	0.02
$m_{\tau\tau,\text{j}_0}$	–	0.57
$E_{\text{T}}^{\text{miss}}/p_{\text{T}}^{\ell_1}$	–	0.09
Sphericity	–	0.08
$m_{\text{T}}^{\ell_0}$	–	0.95
η_{ℓ_0}	–	0.78
$\min \Delta\eta(\ell\ell, \text{jets})$	–	0.03

Table 9.12.: χ^2-test probabilities between the background distributions in the control regions for the VBF category of simulated events and data distributions for the observables used as input variables in the VBF SF and DF regions. Only statistical uncertainties are considered.

variable	χ^2 prob. Zll CR	χ^2 prob. Top CR
$m_{\mathrm{MMC}}^{\mathrm{mlm}}$	0.00	0.62
$\min \Delta\eta(\ell\ell, \mathrm{jets})$	0.14	0.95
m_{jj}	0.00	0.31
$\Delta R(\ell, \ell)$	0.55	0.06
n_{jets}	0.00	0.34
$E_{\mathrm{T}}^{\mathrm{miss}}$ Φ centrality	0.19	0.57
$p_{\mathrm{T}}^{\mathrm{total}}$	0.00	0.08
$E_{\mathrm{T}}^{\mathrm{miss}}$	0.00	0.19
$\min \Delta R(\ell_0, \mathrm{jets})$	0.13	0.08
$\min \Delta R(\ell_1, \mathrm{jets})$	0.14	0.18

Table 9.13.: χ^2-test probabilities between the background distributions in the control regions for the boosted category of simulated events and data distributions for the observables used as input variables in the boosted SF and DF regions. Only statistical uncertainties are considered.

variable	χ^2 prob. Zll CR	χ^2 prob. Top CR
$m_{\mathrm{MMC}}^{\mathrm{mlm}}$	0.00	0.16
$m_{\ell\ell}$	0.14	0.47
$E_{\mathrm{T}}^{\mathrm{miss}}$	0.00	0.20
$\Delta R(\ell, \ell)$	0.00	0.06
$m_{\tau\tau,\mathrm{j}_0}$	0.06	0.71
$E_{\mathrm{T}}^{\mathrm{miss}}/p_{\mathrm{T}}^{\ell_1}$	0.00	0.22
sphericity	0.50	0.17
$m_{\mathrm{T}}^{\ell_0}$	0.00	0.13
η_{ℓ_0}	0.13	0.12
$\min \Delta\eta(\ell\ell, \mathrm{jets})$	0.10	0.55

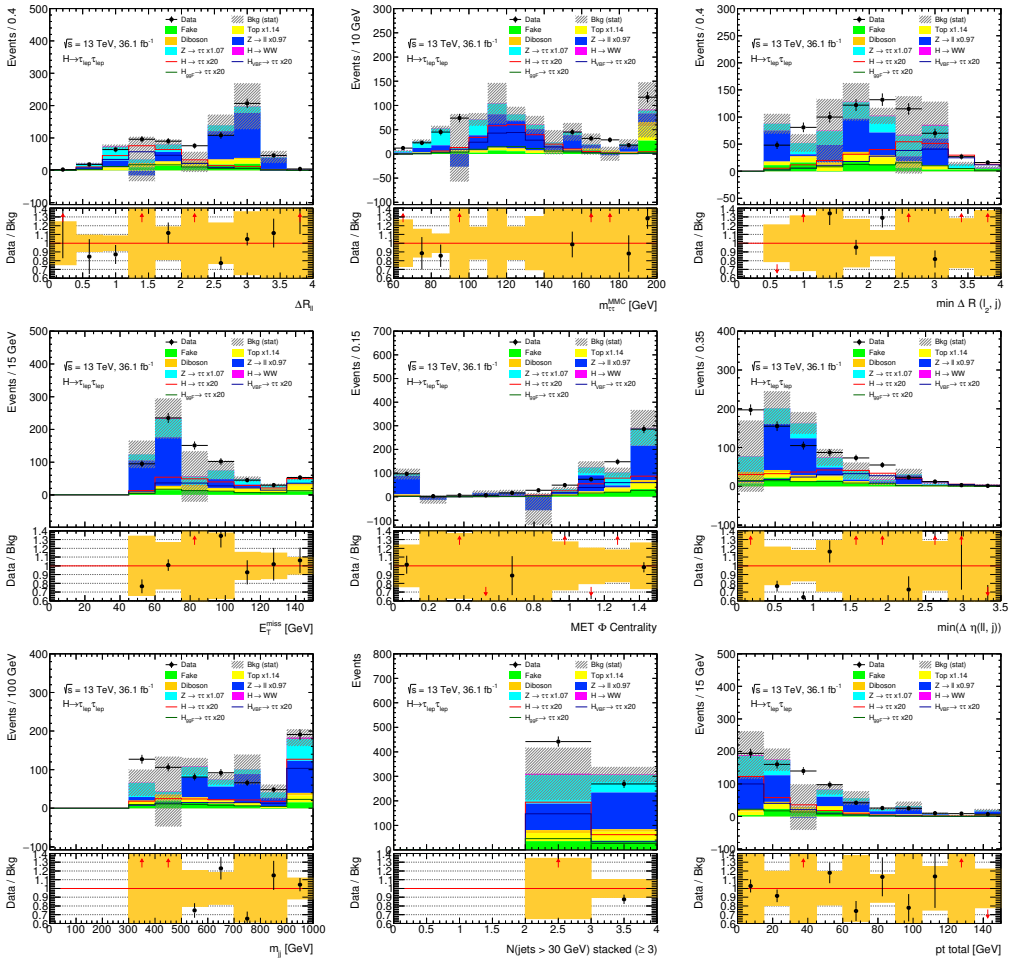


Figure 9.14.: Distributions of the observables which are used as input variables for the BDTs in the VBF SF category. The observables are from top to bottom and from left to right: $\Delta R_{\ell\ell}$, m_{MMC}, $\min \Delta R(\ell_2, \mathrm{jets})$, $E_{\mathrm{T}}^{\mathrm{miss}}$, $E_{\mathrm{T}}^{\mathrm{miss}}\Phi$ centrality, $\min \Delta\eta(\ell\ell, \mathrm{jets})$, m_{jj}, n_{jets}, and $p_{\mathrm{T}}^{\mathrm{total}}$. The signal and background distributions are normalized to their theory cross-sections and luminosity. Additional normalization factors are applied on the top-quark, $Z \to \ell\ell$, and $Z \to \tau\tau$ background. The signal is scaled by a factor of 20. Underflow and overflow bins are included in the first and last bin, respectively. Only statistical uncertainties are included in the error band.

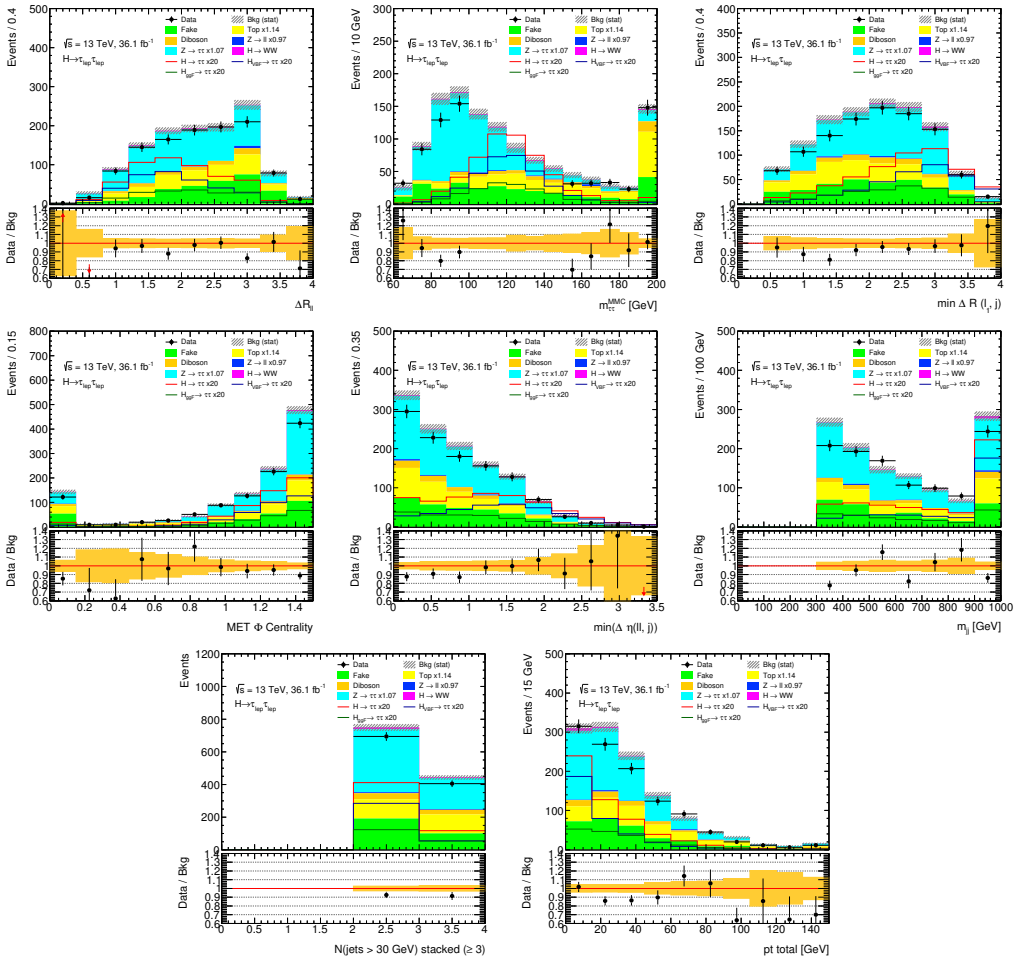


Figure 9.15.: Distributions of the observables which are used as input variables for the BDTs in the VBF DF category. The observables are from top to bottom and from left to right: $\Delta R_{\ell\ell}$, m_{MMC}, $\min \Delta R(\ell_1, \mathrm{jets})$, $E_{\mathrm{T}}^{\mathrm{miss}}\Phi$ centrality, $\min \Delta\eta(\ell\ell, \mathrm{jets})$, m_{jj}, n_{jets}, and $p_{\mathrm{T}}^{\mathrm{total}}$. The signal and background distributions are normalized to their theory cross-sections and luminosity. Additional normalization factors are applied on the top-quark, $Z \to \ell\ell$, and $Z \to \tau\tau$ background. The signal is scaled by a factor of 20. Underflow and overflow bins are included in the first and last bin, respectively. Only statistical uncertainties are included in the error band.

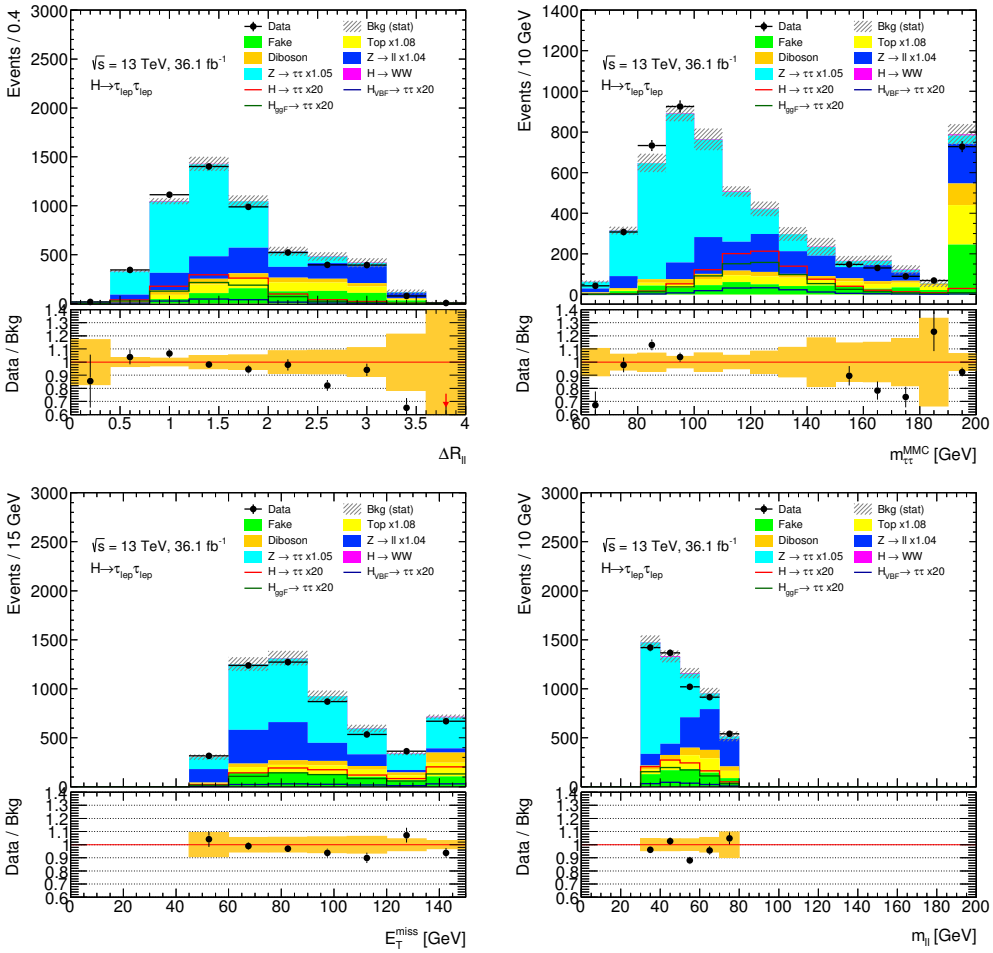


Figure 9.16.: Distributions of the observables which are used as input variables for the BDTs in the boosted SF category. The observables are from top to bottom and from left to right: $\Delta R_{\ell\ell}$, m_{MMC}, $E_{\mathrm{T}}^{\mathrm{miss}}$, and $m_{\ell\ell}$. The signal and background distributions are normalized to their theory cross-sections and luminosity. Additional normalization factors are applied on the top-quark, $Z \to \ell\ell$, and $Z \to \tau\tau$ background. The signal is scaled by a factor of 20. Underflow and overflow bins are included in the first and last bin, respectively. Only statistical uncertainties are included in the error band.

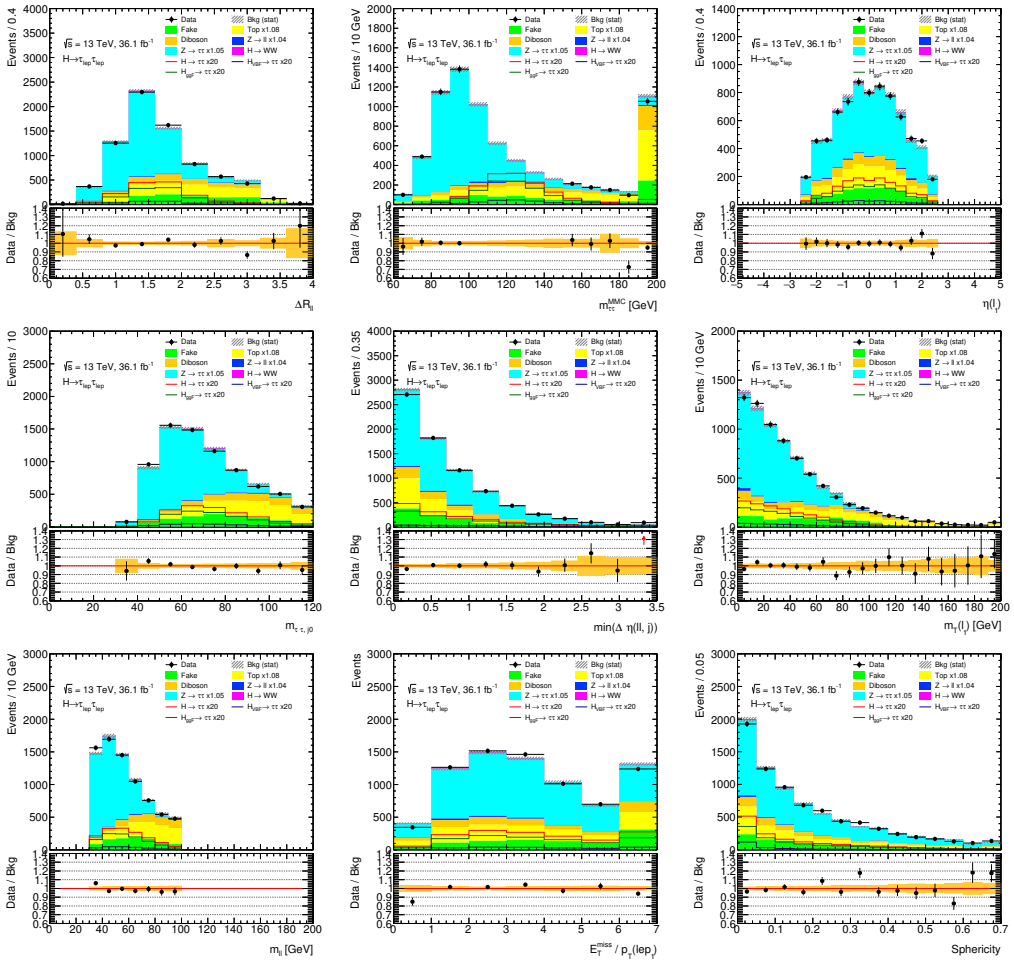


Figure 9.17.: Distributions of the observables which are used as input variables for the BDTs in the boosted DF category. The observables are from top to bottom and from left to right: $\Delta R_{\ell\ell}$, m_{MMC}, η_{ℓ_1}, $m_{\tau\tau,j_1}$, $\min \Delta\eta(\ell, \mathrm{jets})$, $m_{\mathrm{T}}^{\ell_1}$, $m_{\ell\ell}$, $E_{\mathrm{T}}^{\mathrm{miss}}/p_{\mathrm{T}}^{\ell_2}$, and Sphericity. The signal and background distributions are normalized to their theory cross-sections and luminosity. Additional normalization factors are applied on the top-quark, $Z \to \ell\ell$, and $Z \to \tau\tau$ background. The signal is scaled by a factor of 20. Underflow and overflow bins are included in the first and last bin, respectively. Only statistical uncertainties are included in the error band.

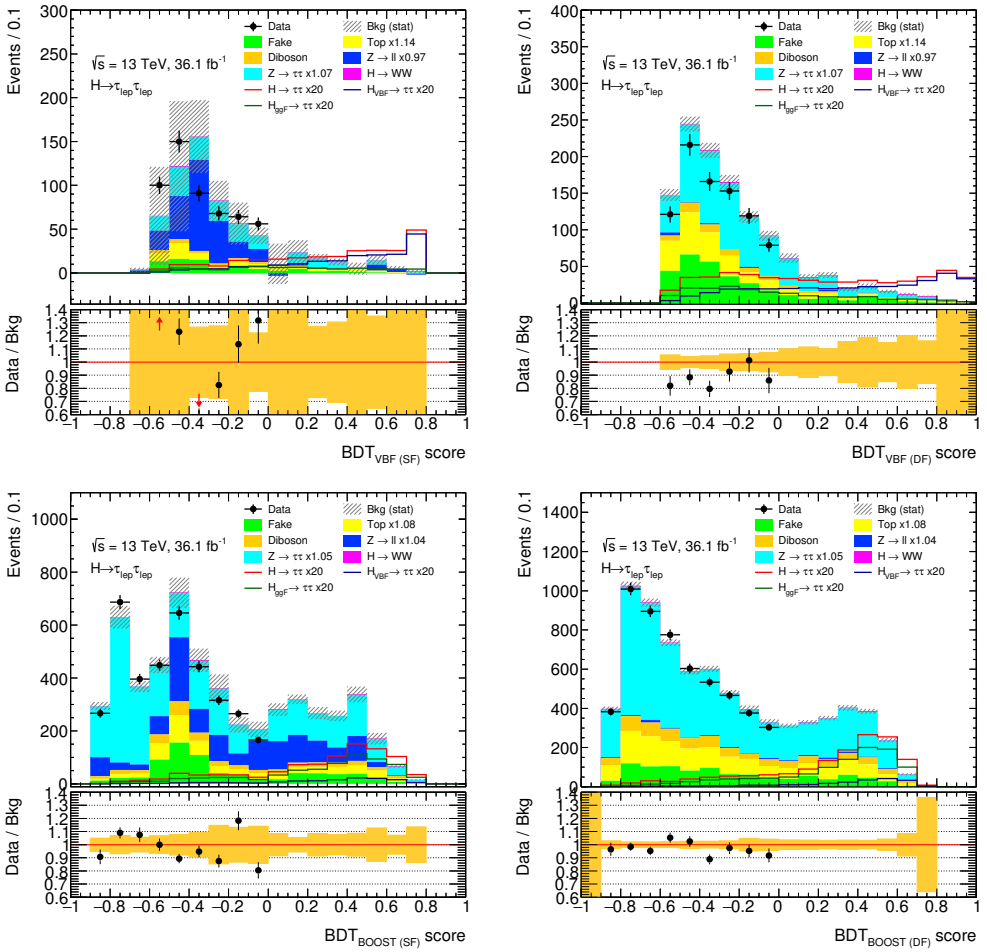


Figure 9.18.: Distributions of BDT outputs in the VBF SF (top left), VBF DF (top right), boosted SF (bottom left), and boosted DF (bottom right) region. The signal and background distributions are normalized to their theory cross-sections and luminosity. Additional normalization factors are applied on the top-quark, $Z \to \ell\ell$, and $Z \to \tau\tau$ background. The signal is scaled by a factor of 20. Underflow and overflow bins are included in the first and last bin, respectively. Only statistical uncertainties are included in the error band.

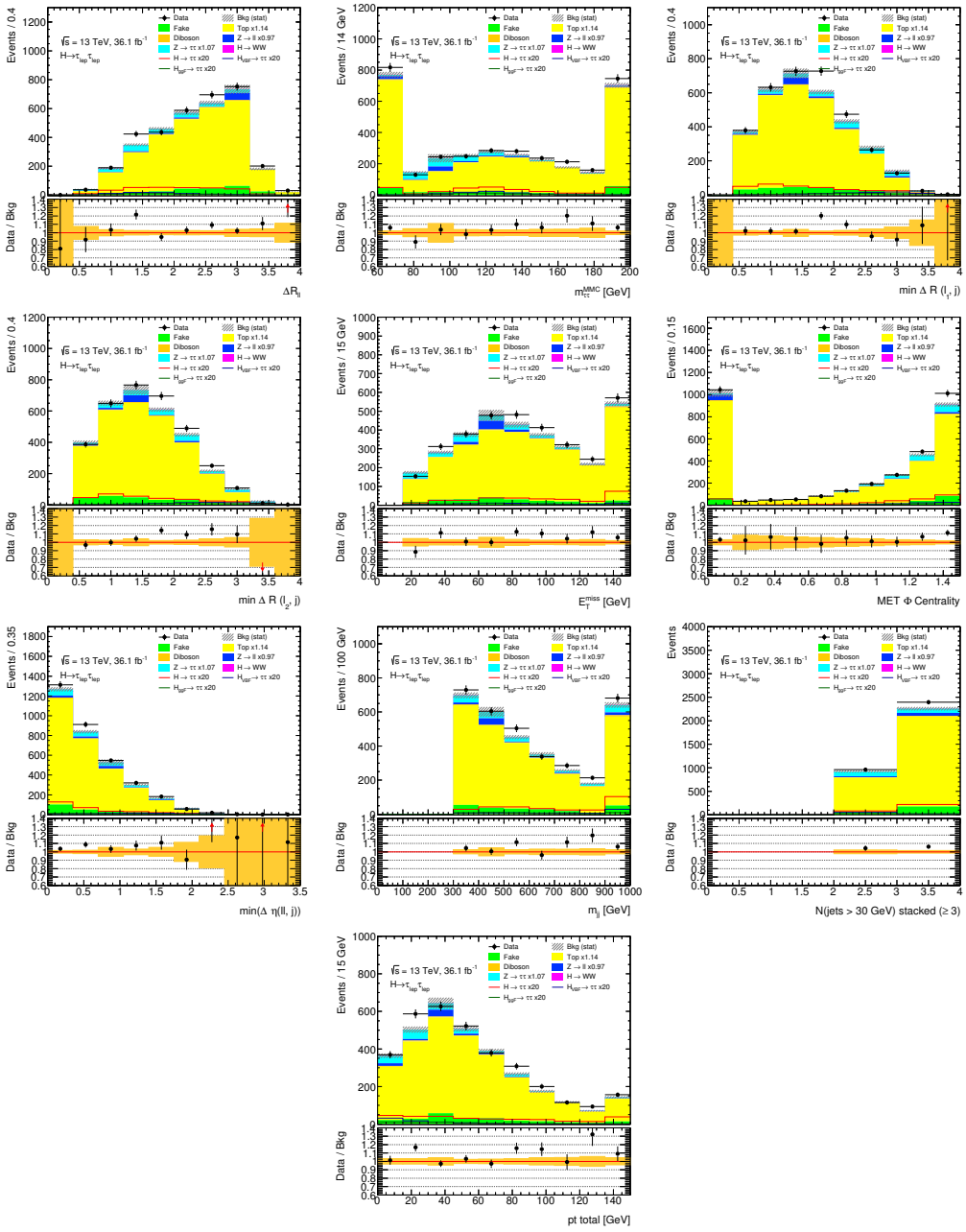


Figure 9.19.: Distributions of the observables which are used as input variables for the VBF SF and VBF DF BDTs in the top-quark control region. The observables are from top to bottom and from left to right: $\Delta R_{\ell\ell}$, m_{MMC}, $\min \Delta R(\ell_1, \mathrm{jets})$, $\min \Delta R(\ell_2, \mathrm{jets})$, $E_{\mathrm{T}}^{\mathrm{miss}}$, $E_{\mathrm{T}}^{\mathrm{miss}}\Phi$ centrality, $\min \Delta\eta(\ell\ell, \mathrm{jets})$, m_{jj}, n_{jets}, and $p_{\mathrm{T}}^{\mathrm{total}}$. The signal and background distributions are normalized to their theory cross-sections and luminosity. Additional normalization factors are applied on the top-quark, $Z \to \ell\ell$, and $Z \to \tau\tau$ background. The signal is scaled by a factor of 20. Underflow and overflow bins are included in the first and last bin, respectively. Only statistical uncertainties are included in the error band.

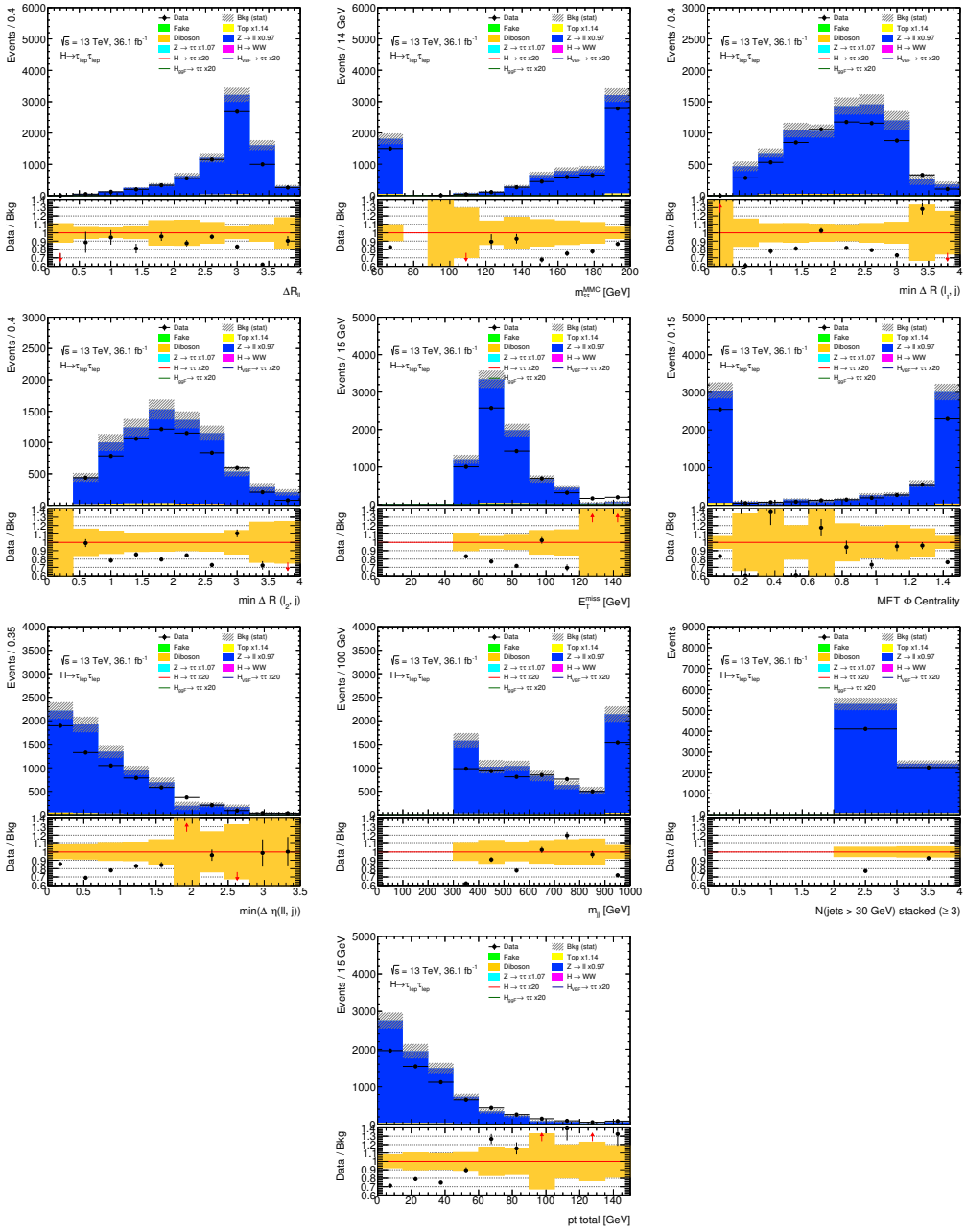


Figure 9.20.: Distributions of the observables which are used as input variables for the VBF SF and VBF DF BDTs in the $Z \to \ell\ell$ control region. The observables are from top to bottom and from left to right: $\Delta R_{\ell\ell}$, m_{MMC}, $\min \Delta R(\ell_1, \mathrm{jets})$, $\min \Delta R(\ell_2, \mathrm{jets})$, $E_{\mathrm{T}}^{\mathrm{miss}}$, $E_{\mathrm{T}}^{\mathrm{miss}}\Phi$ centrality, $\min \Delta\eta(\ell\ell, \mathrm{jets})$, m_{jj}, n_{jets}, and $p_{\mathrm{T}}^{\mathrm{total}}$. The signal and background distributions are normalized to their theory cross-sections and luminosity. Additional normalization factors are applied on the top-quark, $Z \to \ell\ell$, and $Z \to \tau\tau$ background. The signal is scaled by a factor of 20. Underflow and overflow bins are included in the first and last bin, respectively. Only statistical uncertainties are included in the error band.

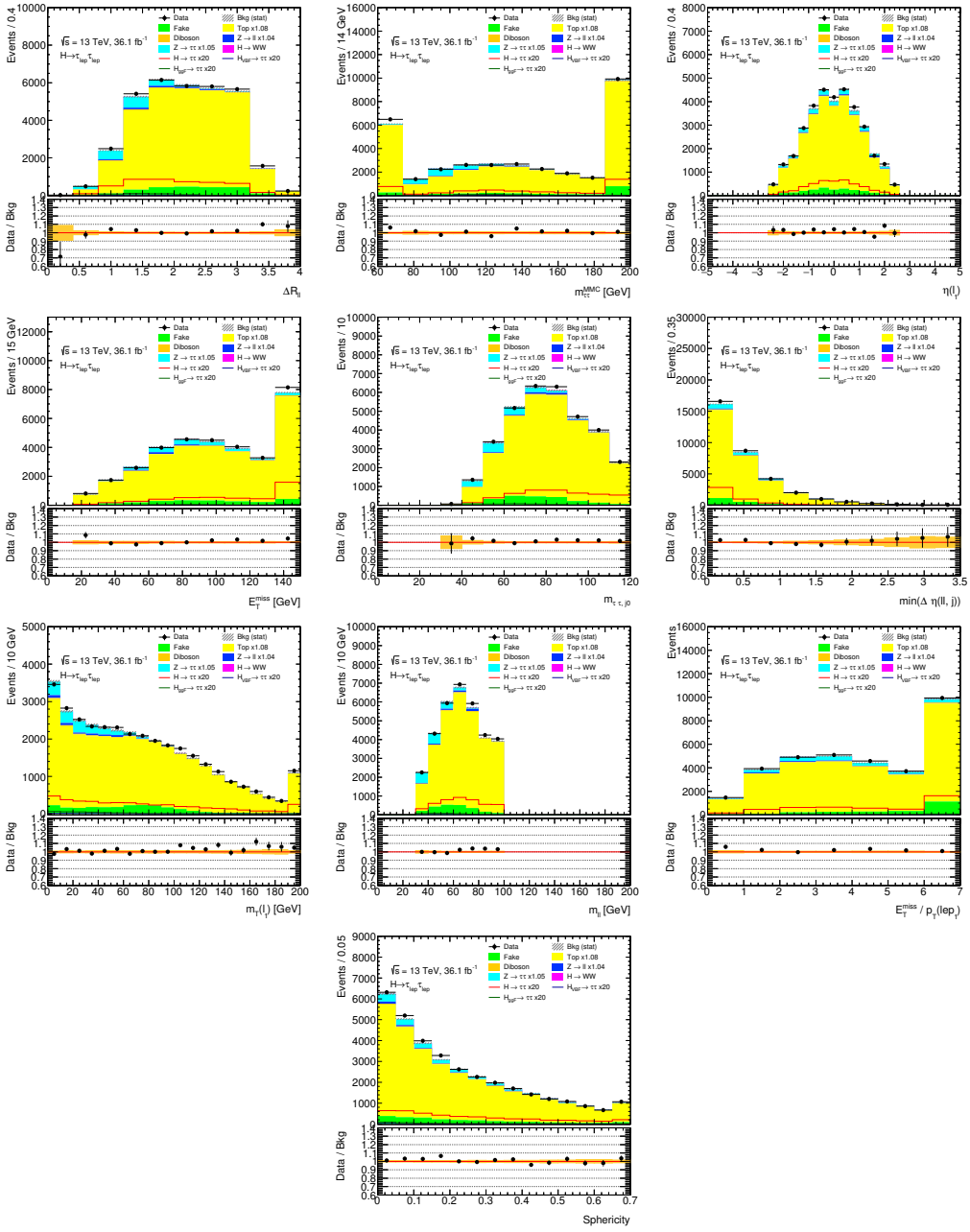


Figure 9.21.: Distributions of the observables which are used as input variables for the boosted SF and boosted DF BDTs in the top-quark control region. The observables are from top to bottom and from left to right: $\Delta R_{\ell\ell}$, m_{MMC}, η_{ℓ_1}, $E_{\mathrm{T}}^{\mathrm{miss}}$, $m_{\tau\tau,j_1}$, $\min \Delta\eta(\ell, \mathrm{jets})$, $m_{\mathrm{T}}^{\ell_1}$, $m_{\ell\ell}$, $E_{\mathrm{T}}^{\mathrm{miss}}/p_{\mathrm{T}}^{\ell_2}$, and Sphericity. The signal and background distributions are normalized to their theory cross-sections and luminosity. Additional normalization factors are applied on the top-quark, $Z \to \ell\ell$, and $Z \to \tau\tau$ background. The signal is scaled by a factor of 20. Underflow and overflow bins are included in the first and last bin, respectively. Only statistical uncertainties are included in the error band.

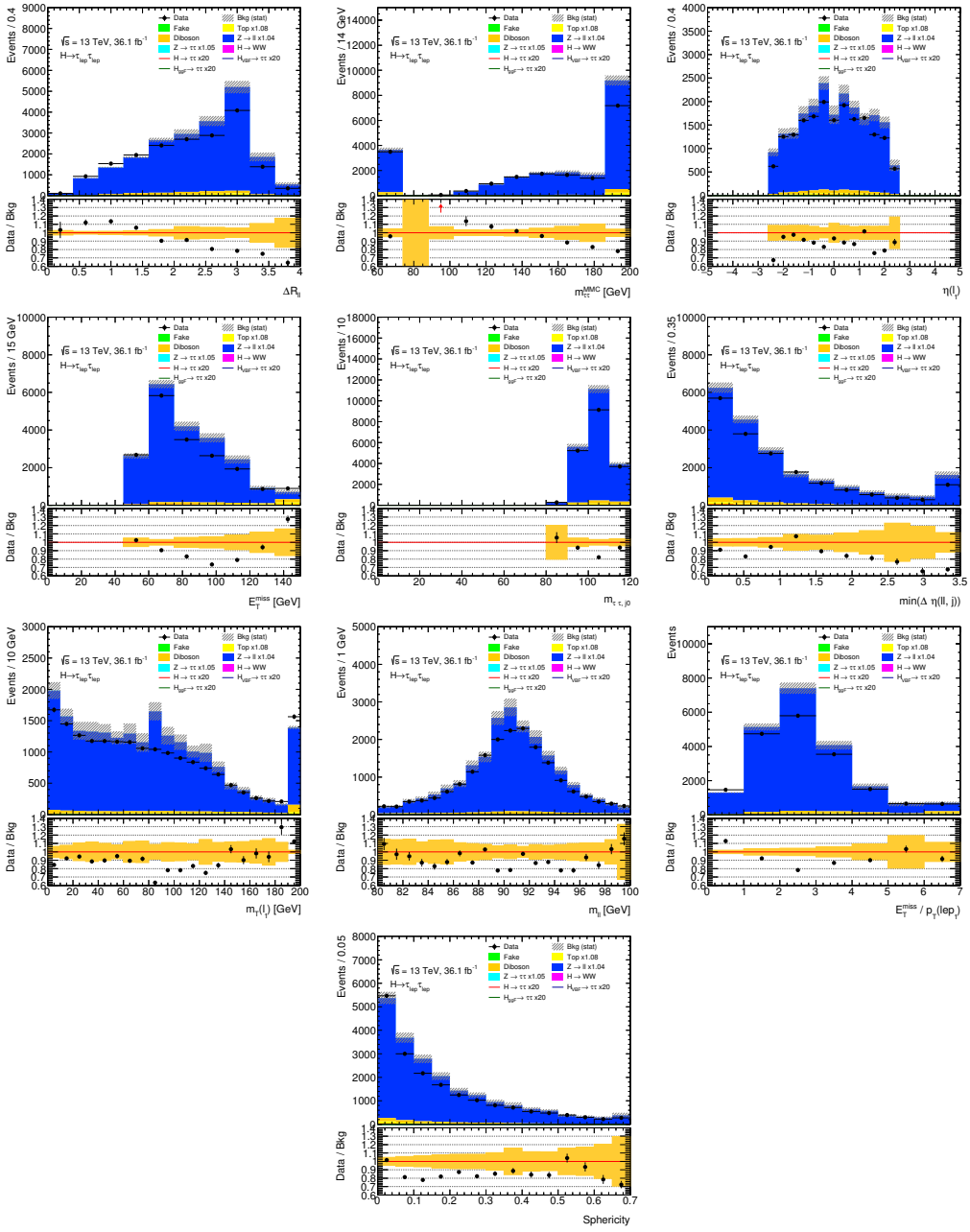


Figure 9.22.: Distributions of the observables which are used as input variables for the boosted SF and boosted DF BDTs in the $Z \to \ell\ell$ control region. The observables are from top to bottom and from left to right: $\Delta R_{\ell\ell}$, m_{MMC}, η_{ℓ_1}, $E_{\mathrm{T}}^{\mathrm{miss}}$, $m_{\tau\tau,j_1}$, $\min \Delta\eta(\ell, \mathrm{jets})$, $m_{\mathrm{T}}^{\ell_1}$, $m_{\ell\ell}$, $E_{\mathrm{T}}^{\mathrm{miss}}/p_{\mathrm{T}}^{\ell_2}$, and Sphericity. The signal and background distributions are normalized to their theory cross-sections and luminosity. Additional normalization factors are applied on the top-quark, $Z \to \ell\ell$, and $Z \to \tau\tau$ background. The signal is scaled by a factor of 20. Underflow and overflow bins are included in the first and last bin, respectively. Only statistical uncertainties are included in the error band.

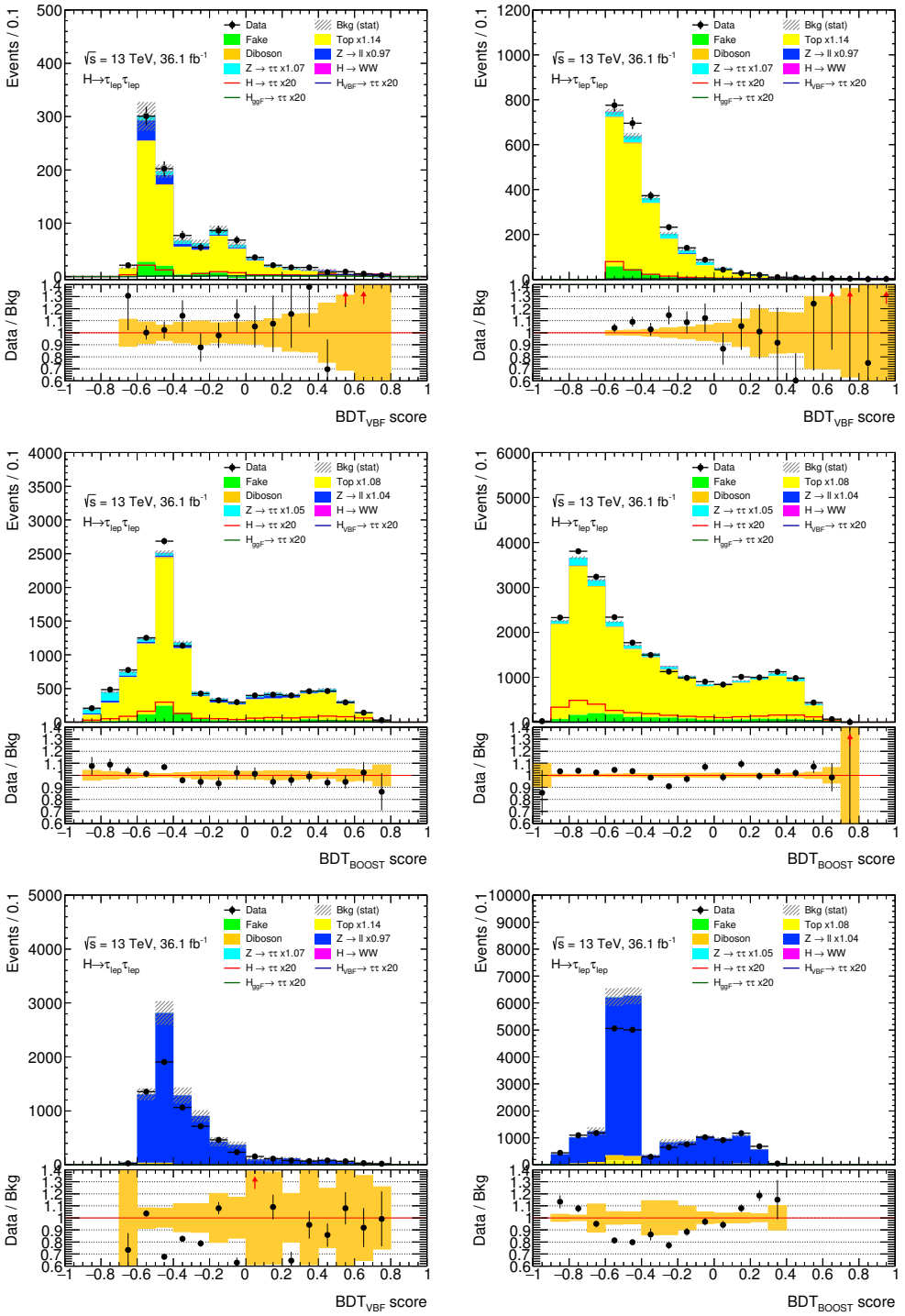


Figure 9.23.: Distributions of BDT outputs in the top-quark control region (VBF SF: top left, VBF DF: top right, boosted SF: center left, boosted DF: center right) and $Z \to \ell\ell$ control region (VBF SF: bottom left, boosted SF bottom right). The signal and background distributions are normalized to their theory cross-sections and luminosity. Additional normalization factors are applied on the top-quark, $Z \to \ell\ell$, and $Z \to \tau\tau$ background. The signal is scaled by a factor of 20. Underflow and overflow bins are included in the first and last bin, respectively. Only statistical uncertainties are included in the error band.

10. Systematic Uncertainties

Systematic uncertainties arise from various sources in the analysis of $H \to \tau^+\tau^- \to \ell^+\ell^- 4\nu$. They affect the event yields in the final distributions, which are used in the fit for extracting the signal strength. There are two types of systematic uncertainties: the *shape uncertainties*, which have an impact on the distributions of the used observables, and the *normalization systematics*, which affect the expected signal and background yields.
Systematic uncertainties can by classified by the type of their source. Experimental systematics arise due to measurements of for example the luminosity or efficiencies of object identification and reconstruction. The calibration of the detectors, especially the calorimeters, is another contribution. Another source are uncertainties related to theory predictions. These depend for example on the choice of the renormalization and factorization scales and uncertainties in α_s and PDFs. Additionally, the data-driven estimation of background processes result in another class of systematic uncertainties.
For most of the systematic uncertainties the source is varied upwards or downwards within one standard deviation and the full analysis is repeated on the modified input, in order to propagate the systematic uncertainty correctly. Then, the impact on the shape of the observable used in the fit and the total signal and background yield is compared with respect to the nominal case. Systematic variations which have only a very minor impact are discarded by a *pruning* procedure.
This chapter gives an overview of the systematic uncertainties which are considered in this analysis. Their incorporation in the fit is discussed in Chapter 11.

10.1. Experimental uncertainties

Luminosity: The uncertainty of the integrated luminosity is $\pm 2.1\,\%$ for the combined 2015 and 2016 dataset [74]. This value was derived in a procedure similar to the method described in [136] from a calibration of the luminosity scale using x-y beam-separation scans performed in August 2015 and May 2016.

Electrons: The uncertainties on the electron trigger, identification, and isolation efficiencies are derived in $J/\Psi \to ee$ and $Z \to ee$ events. The relative variations are within $0.5\,\%$ and $5\,\%$ [104, 137]. Additionally, systematic variations on the electron energy resolution are applied, which is most of the time less than $1\,\%$ [138].

Muons: The muon momentum scale and energy resolutions are varied by $\pm 1\sigma$ in the event reconstruction. The variations are derived in $J/\Psi \to \mu\mu$ and $Z \to \mu\mu$ are range between $1.7\,\%$ and $2.9\,\%$. The systematic variations on the trigger,

identification, and isolation efficiencies for muons are derived in the same events. The variations are between 1 % and 7 % [105].

Jet energy: The jet uncertainties depend on the transverse energy and η of the jet. The uncertainties on the jet energy resolution (JER) and jet energy scale (JES) are obtained by smearing the nominal jet energy resolution. The JES uncertainty is 6 % for jets with $p_{\mathrm{T}} = 20\,\mathrm{GeV}$, decreases to 1 % for jets between with a transverse momentum between 200 GeV and 1800 GeV and rises again to 3 % for jets with higher p_{T}. For the JER and JES uncertainties the 11 and 19 parameter scheme is used, respectively. [108, 139]

Jet vertex tagger: The jet vertex tagger (JVT) algorithm is used to suppress jets from pile-up events by applying a threshold on the output of the JVT algorithm, as described in Section 5.4. The variation of this threshold is used as a systematic uncertainty.

b-tagging: Scale factors are used to correct the b-tagging efficiencies and mistag rate for light flavour jets. This class of systematic variations refers to the variation of these scale factors. The systematic variation affects mainly the single-top and $t\bar{t}$ background.

Transverse missing energy: For the systematic uncertainty on the missing transverse energy the different contribution is smeared by a convolution of a Gaussian and fitted to data. The parameters of the Gaussian are used to determine the uncertainties of $E_{\mathrm{T}}^{\mathrm{miss}}$. [140].

Pile-up reweighting: In the generation of simulated events a generalized profile for the distribution of the number of interactions per bunch crossing (pile-up distribution) is used. To match with the observed profile of the 2015 and 2016 dataset, a correction factor of 1/1.16 needs to be applied to the pile-up distribution.. It was determined that the uncertainty for a 1σ variation is $1/(1.16 \pm 0.07)$. However, the more conservative estimation of $1/(1.16^{+0.07}_{-0.16})$ is used in this analysis.

10.2. Uncertainties on data-driven background estimations

The background of events with misidentified leptons (fake leptons) is estimated in this analysis in a data-driven way, as described in Section 7.1. The systematic variations are constructed from variations of the efficiencies of the triggers, which are used to select the events. Additionally, non-closure systematics are derived by comparing data events from the opposite sign and same sign fake control region. An overview of the variations in the non-closure uncertainties is given in Table 10.1.

Table 10.1.: Overview of the non-closure systematic variations for the fake background for different bins of the transverse momentum of the subleading lepton, $p_{\mathrm{T}}^{\ell_2}$.

$p_{\mathrm{T}}^{\ell_2}$ [GeV]	VBF cat.	Boosted cat.
$[10, 15)$	39 %	37 %
$[15, 20)$	90 %	42 %
$[20, 25)$	90 %	24 %
$[25, \infty)$	90 %	24 %

10.3. Theory uncertainties

Signal theory systmatics: For the gluon–gluon fusion, vector boson fusion, and Higgs-Strahlung production modes of the Higgs boson several systematic variations are applied. The sources of these variations are missing higher order calculations, which also affect the uncertainty of jet bin migration, the underlying event and hadronization, and the uncertainty on the PDFs and α_s.

The event yield in each bin of the final distribution which is used in the fit can be expressed by

$$n_{\mathrm{bin}} = \sigma_{\mathrm{tot}} \times \mathcal{L} \times A \times f_{\mathrm{bin}} \,, \tag{10.1}$$

with the total cross section σ_{tot}, the luminosity $\mathcal{L}$, the acceptance A of the total considered sample in the signal region, and the fraction of events in the considered bin, f_{bin}. The uncertainties are factorized into contributions on the total cross section, acceptance, and the shape of the distributions.

The uncertainties on the total cross section of the production modes are shown in Table 2.4.

Uncertainties due to missing higher orders are estimated by scaling the factorization and renormalization scales by a factor of 0.5 and 2. The total up and down uncertainties the VBF production mode are estimated by the maximum negative and positive variations in the acceptances. For the gluon–gluon fusion production mode jet bin migration uncertainties need to be considered. Here, a method using the $H + 0$jet, $H + 1$jet, and $H+ \geq 2$jets cross sections recommended by [46] is applied, where 4 scale variations, 2 VBF topology uncertainties, 2 Higgs p_{T}-shape uncertainties, and one top mass dependence are considered.

The uncertainties on the underlying event are studied on truth level by replacing the parton shower algorithm with a different one (`Herwig7` [141, 142]). This is a normalization uncertainty in the acceptance.

For the PDF uncertainties the latest PDF4LHC15 [143] recommendations are used, which are based on 30 eigenvector variations and 2 variations for α_s.

$Z\tau\tau$ theory systmatics: For the $Z \to \tau\tau$ signal uncertainties scale variations and PDF uncertainties are considered. The scale uncertainties are evaluated the same

way as the signal VBF uncertainties. For the PDF uncertainties the same scheme as for the signal is used.

$H \to \tau\tau$ branching ratio: The uncertainty on the $H \to \tau\tau$ branching ratio is $\pm 5.7\,\%$ [47].

11. Statistical Analysis and Results

In this chapter the likelihood fit is introduced, which is used to determine the signal strength of the $H \to \tau^+\tau^- \to \ell^+\ell^- 4\nu$ process. For the multivariate analysis (MVA) the shape of the BDT output is used as the final discriminant. In the cut-based analysis (CBA) the m_{MMC} distribution is used. Furthermore, a strategy to obtain an optimal binning for the distribution of the BDT output is discussed. Finally, the fit results are presented and compared between MVA and CBA.

11.1. Likelihood fit

The likelihood fit is used to determine the signal strength μ, which is the ratio of the measured (fitted) signal cross-section to the Standard Model prediction for cross-section times branching ratio of $H \to \tau_{\text{lep}}\tau_{\text{lep}}$. A value of $\mu = 0$ indicates the absence of signal, $\mu = 1$ corresponds to the case that the amount of signal matches with the SM prediction. The fit is executed with the `HistFactory` tool [144] using the `RooStats` [145] package and the `WSMaker` [146] script collection.
The fit is based on a binned likelihood function $\mathcal{L}$ [147], which originates from the Poisson distribution,

$$\mathcal{L}(\mu, \boldsymbol{\theta}) = \prod_{i=1}^{N_{\text{bins}}} = \frac{(\mu s_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta}))^{n_i} e^{-(\mu s_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta}))}}{n_i!} \,. \tag{11.1}$$

The binned likelihood function gives the probability of finding n data events in the bin i, where s signal events and b background events are expected. Here, the bins are not limited to one observable, but different contributions, for example single bin categories and binned distributions, can be combined in the likelihood. Additionally, the likelihood function depends on several nuisance parameters, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$, which are discussed below. A large value of the likelihood function indicates a better agreement between prediction and data.
The best fit value for μ is obtained by maximizing the likelihood function, and is denoted as $\hat{\mu}$. However, it is more convenient to minimize the negative logarithm of the likelihood function, $-\ln \mathcal{L}$, which is also known as the negative log-likelihood (NLL).
In the search of new physics the compatibility between the observed data and the background-only hypothesis is an important quantity. Therefore, the null hypothesis is $\mu = 0$ which needs to be rejected for an observation of the signal process.

First, a test statistic q_0 is constructed [147],

$$q_0 = \begin{cases} -2\ln(\mathcal{L}(0,\hat{\hat{\boldsymbol{\theta}}})/\mathcal{L}(\hat{\mu},\hat{\theta})) & \hat{\mu} \geq 0 \\ 0 & \hat{\mu} < 0 \end{cases}, \tag{11.2}$$

where $\mathcal{L}(\hat{\mu},\hat{\theta})$ refers to the maximum of the likelihood function and $\mathcal{L}(0,\hat{\hat{\boldsymbol{\theta}}})$ to an conditional maximum where μ is fixed according to the null hypothesis. The test statistic should have a good separation between the null hypothesis and the alternative hypothesis, where the signal strength is non-vanishing. If a negative amount of signal events is observed ($\hat{\mu} < 0$), which is unphysical, the test statistic is set to zero, since a positive amount is expected. Using the probability density function of the test statistic, $g(q_0|0,\hat{\hat{\boldsymbol{\theta}}})$, the compatibility of the null hypothesis with the observed test statistic $q_{0,\text{obs}}$ can be calculated, which is expressed with the p_0 value,

$$p_0 = \int_{q_0,\text{obs}}^{\infty} g(q_0|0,\hat{\hat{\boldsymbol{\theta}}})\,\mathrm{d}q_0\,. \tag{11.3}$$

The p_0 value is the probability that the test statistic q_0 is at least as incompatible with the null hypothesis (here: background only) as the observed one, assuming that the background only hypothesis is realized in nature.
Usually the p_0-value is expressed in terms of Gaussian standard deviations. This is the so-called *significance* Z_0, which can be calculated by

$$Z_0 = \Phi^{-1}(1-p_0) \simeq \sqrt{q_0}\,. \tag{11.4}$$

Here, Φ is the cumulative distribution function of the Gaussian distribution. The asymptotic approximation [147] is used in Eq. (11.4) to simplify the calculation.
An observation is claimed for a p_0-value of $p_0 < 2.85 \times 10^{-7}$, which corresponds to a significance of $Z_0 \geq 5$.

11.2. Fitting Procedure

For the construction of the likelihood function several signal and control regions are used, which are discussed below.
In the cut-based analysis four signal regions, two for the VBF category and two for the boosted category as explained in Section 6.4 is used. The binning is setup in such a ways that there is at least some amount of background in each bin. Under and overflow bins are included in the first and last bin, respectively.
An overview of the binning is given in Table 11.1. To obtain the normalization of the $Z \to \ell\ell$and top-background a dedicated control region is used for each background. There are separate control regions for the VBF and boosted category. The definitions of the control regions can be found in Section 7.2. Since the control regions are used only for normalization purposes, only a single bin is used, which contains the complete event count.

Table 11.1.: Overview of the binning in the signal regions (SR) and control regions (CR) used in the fit for the cut-based analysis.

Region	Type	Binning [GeV]
high VBF	SR	$[60, 88, 116, 130, 200]$
low VBF	SR	$[60, 88, 116, 130, 200]$
high boosted	SR	$[60, 80, 110, 120, 130, 140, 200]$
low boosted	SR	$[60, 80, 110, 120, 130, 140, 200]$
$Z \to \ell\ell$ VBF control region (CBA)	CR	single bin
Top VBF control region (CBA)	CR	single bin
$Z \to \ell\ell$ boosted control region (CBA)	CR	single bin
Top boosted control region (CBA)	CR	single bin

In the multivariate analysis a similar approach is applied. For the signal regions distributions of the BDT output after the modified event selection (Section 9.1) are used. In the boosted category the individual BDTs for same flavour (SF) and different flavour (DF) events are taken. However, it turned out that splitting the events into SF and DF in the VBF category actually decreased the significance, because insufficient amount of statistics. Therefore, the BDTs for SF and DF events are combined by adding up the event yields in the individual bins. The control regions are based on the modified event selection, but no other changes were made.
The binning for the BDT distributions is calculated with a method similar to the one used in the Run-1 analysis of the $H \to \tau_{\text{lep}}\tau_{\text{had}}$ channel [148]. Most signal events can be found at higher values of the BDT output distribution, whereas most background events are located at low values. A fine binning would allow for a better measurement of the signal strength, since bins in the region of high BDT values would contain only a small number of background events. However, each bin must contain at least some amount of background events to ensure that statistical fluctuations of the background do not lead to empty bins. The following algorithm provides the finest binning possible without violating the robustness of the background estimation. A fine binned BDT distribution with 200 bins is used as an input.
Starting from the very signal like end of the BDT distribution the bins are merged if all of the following criteria are met:

- The expected yield of the total background has to be larger than the expected yield of total background events in the current bin.

- The relative statistical uncertainty of each background ($Z \to \ell\ell$, $Z \to \tau\tau$, diboson, top, fake, $H \to WW$) has to be smaller than 100 %.

- Either the expected yield yield of the total background is 50 % larger than in the previous bin or the signal-to-background ratio is at most 10 % smaller than in the previous bin.

The first criterion ensures a background shape which is monotonically decreasing. Due to the second criterion it is ensured that statistical fluctuations of the background do not lead to empty bins. The third criterion is used to create a new bin, if the bin offers higher signal sensitivity or if the background composition is much different than in the previous bin.
The algorithm is applied on the BDTs for SF and DF events in the boosted category and on the combined BDT in the VBF category. An overview of the different regions and binnings used in the multivariate analysis is shown in Table 11.2.

Table 11.2.: Overview of the binning in the signal regions (SR) and control regions (CR) used in the fit for the multivariate analysis.

Region	Type	Binning
VBF	SR	$[-1, -0.47, -0.24, 0.19, 0.59, 0.74, 1]$
Boosted SF	SR	$[-1, -0.78, -0.44, -0.19, 0.12, 0.29, 0.41, 0.49, 0.57, 0.64, 0.69, 0.72, 1]$
Boosted DF	SR	$[-1, -0.78, -0.5, -0.21, 0.1, 0.31, 0.42, 0.5, 0.57, 0.61, 0.65, 0.67, 0.69, 1]$
$Z \to \ell\ell$ VBF control region (MVA)	CR	single bin
Top VBF control region (MVA)	CR	single bin
$Z \to \ell\ell$ boosted control region (MVA)	CR	single bin
Top boosted control region (MVA)	CR	single bin

11.3. Results

The likelihood fit is carried out for both the cut-based and multivariate analysis. Because this analysis is still under development and not yet approved by the ATLAS collaboration, the fit cannot be done with the observed data. Instead, the so-called *Asimov data* [147] are used, where the simulated events act as an replacement for the observed data. Therefore, the extracted signal strength will always be one. But the fit can still be used to obtain an expected sensitivity and to validate the fitting procedure. The distributions of m_{MMC} which are used in the fit of the CBA are shown in Figs. 11.1 and 11.2, whereas the corresponding distributions of the BDT output for the MVA are shown in Figs. 11.3 and 11.4. As can be seen the control regions are relatively pure in their respectively processes, see also Section 7.2. The contributions of the "fake" background in the $Z \to \ell\ell$ control regions are minimal and therefore neglected.
The expected significances for the cut-based and multivariate analysis are shown for fits in the individual signal regions and for the combination of all signal regions in Table 11.3. The multivariate analysis manages to increase the sensitivity by approximately 33 % and 72 % in the boosted and VBF category, respectively. For the combined fit of the boosted and VBF category the increase is about 63 %.

Table 11.3.: Expected significances for the cut-based (CBA) and multivariate (MVA) analysis of $H \to \tau_{\text{lep}}\tau_{\text{lep}}$ for the combined and individual categories in units of σ.

Analysis	Combined	Boosted	Boosted SF	Boosted DF	VBF
CBA	0.83	0.58	–	–	0.54
MVA	1.35	0.77	0.51	0.59	0.93

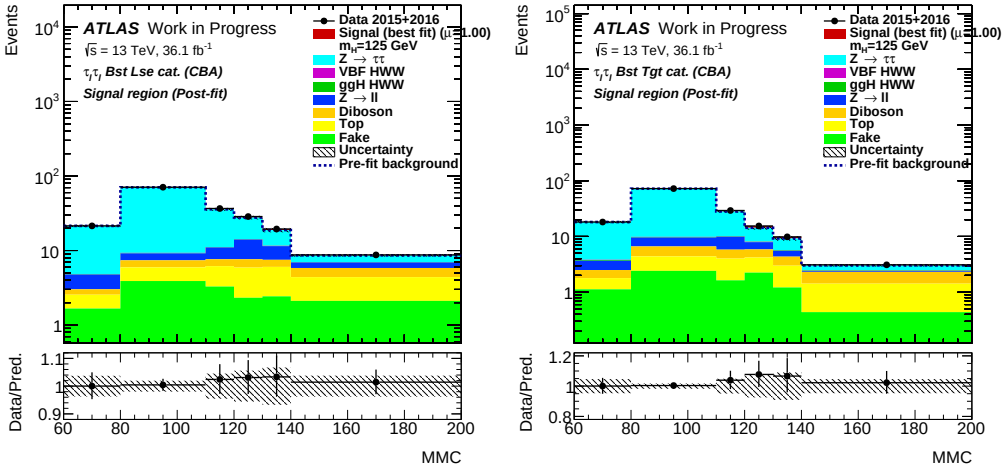
The results of the signal strength measurement in the cut-base and multivariate analysis are

$$\mu_{\text{CBA}} = 1.0\,^{+0.47}_{-0.46}\ (\text{stat.})\ ^{+1.25}_{-1.12}\ (\text{sys.}) \qquad \text{and} \tag{11.5}$$

$$\mu_{\text{MVA}} = 1.0\,^{+0.32}_{-0.31}\ (\text{stat.})\ ^{+0.55}_{-0.66}\ (\text{sys.})\ , \tag{11.6}$$

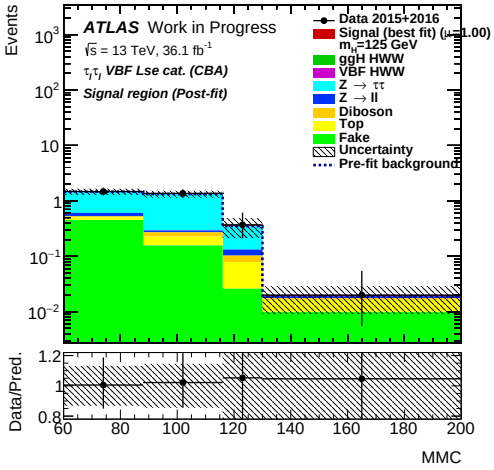
respectively. Additionally, the results are visualized in Fig. 11.5, where also the individual results for the VBF and boosted categories are shown. Since Asimov data are used in the fit, the expected values are always 1. However, it can be seen, that both the statistic and systematic uncertainties on the signal strengths are lower in the MVA than in the CBA. The largest reduction of the uncertainties on the signal strength is in the VBF category, where the uncertainties decrease by a factor between 1.5 and 3. The signal strength of the combined fit and measurement in the boosted category the uncertainties are reduced by a factor of up to two.
A breakdown of the individual contributions to the uncertainties are given in Table 11.4. The most dominate sources of uncertainty are related to measurements of jets and the missing transverse energy. Furthermore, uncertainties due to count

statistics in simulation and normalization of the $Z \to \ell\ell$, $Z \to \tau\tau$, and top-quark background are also major sources of uncertainty.

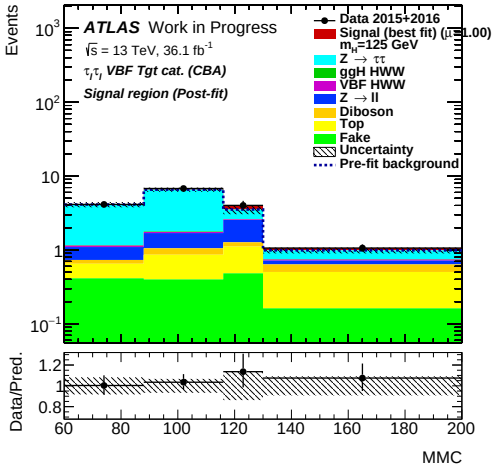


(a) low boosted

(b) high boosted

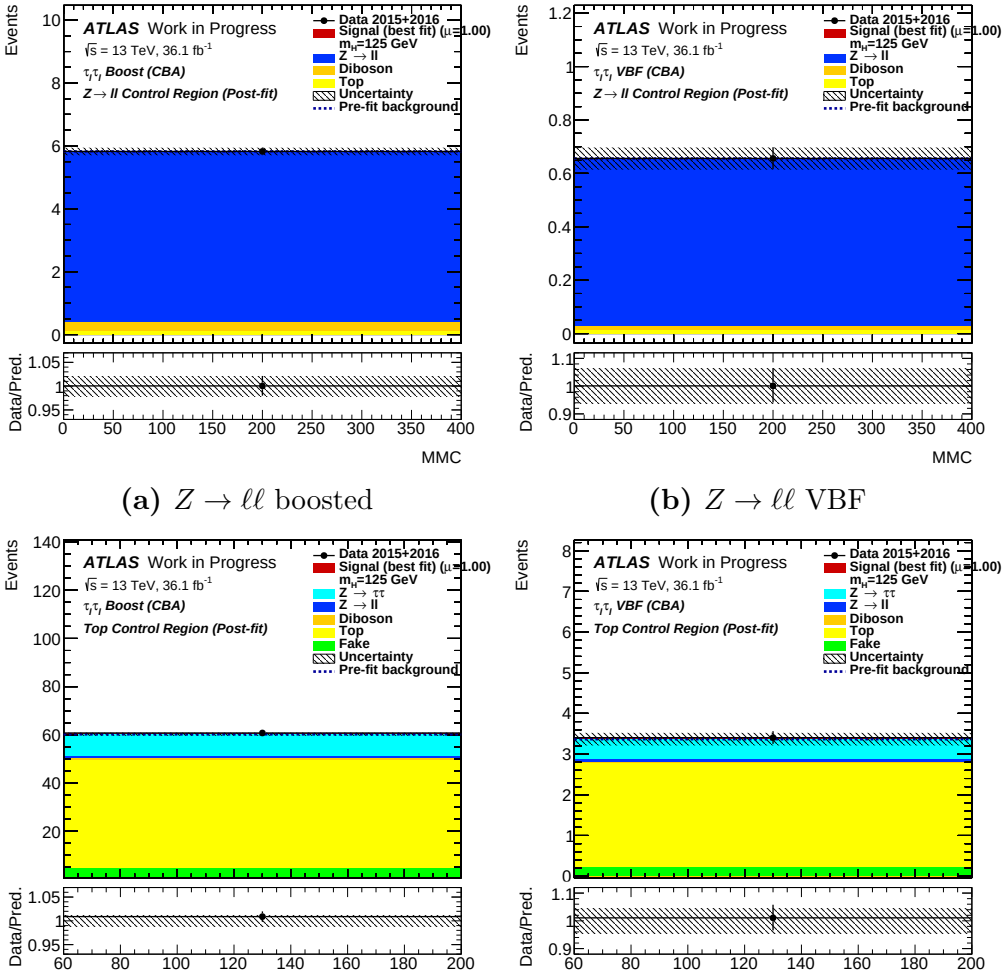


(c) low VBF



(d) high VBF

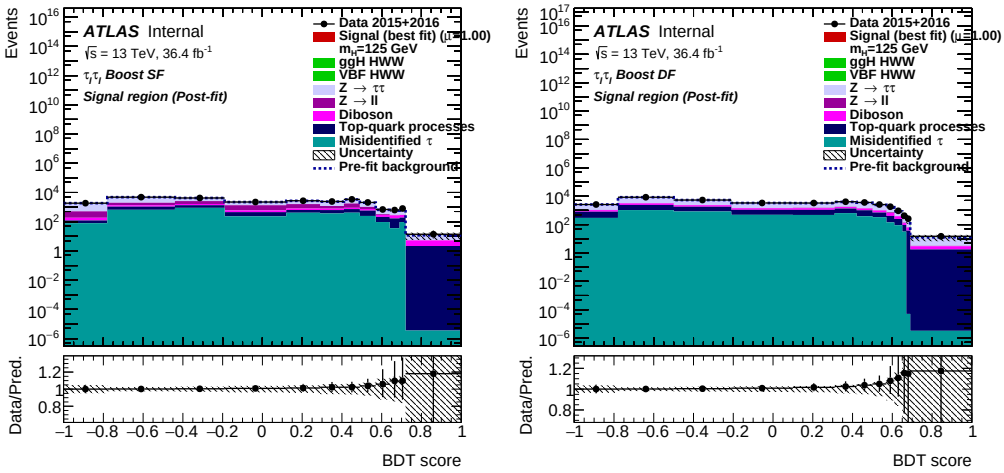
Figure 11.1.: Distributions of m_{MMC} in the different signal regions with the binning used in the fit of the cut-based analysis. For the data shown in the plot Asimov data corresponding to the 2015 and 2016 dataset with 36.1 fb^{-1} are used.



(a) $Z \to \ell\ell$ boosted **(b)** $Z \to \ell\ell$ VBF

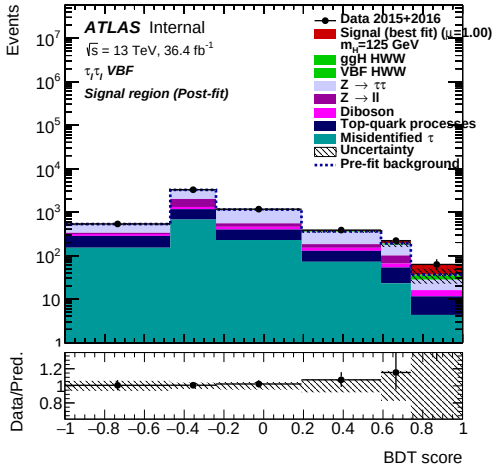
(c) Top boosted **(d)** Top VBF

Figure 11.2.: Distributions of m_{MMC} in the different control regions with the binning used in the fit of the cut-based analysis. For the data shown in the plot Asimov data corresponding to the 2015 and 2016 dataset with 36.1 fb^{-1} are used.



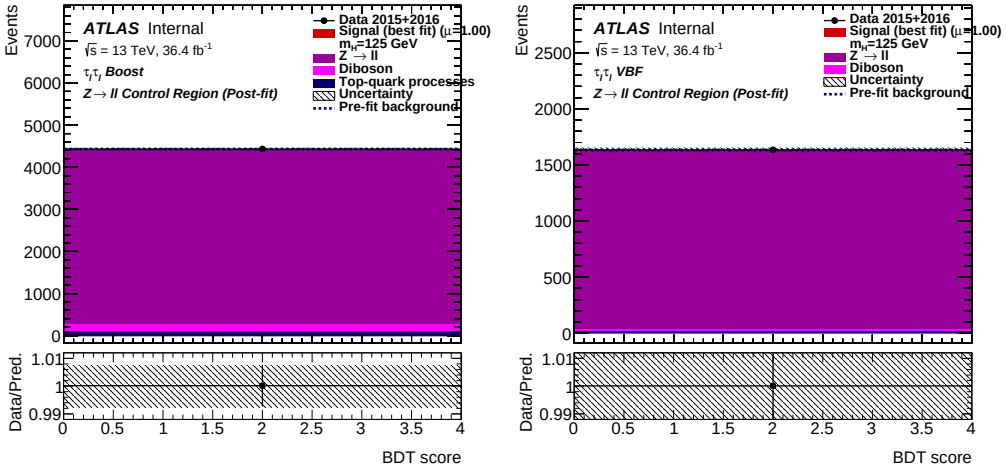
(a) Boosted SF

(b) Boosted DF



(c) VBF

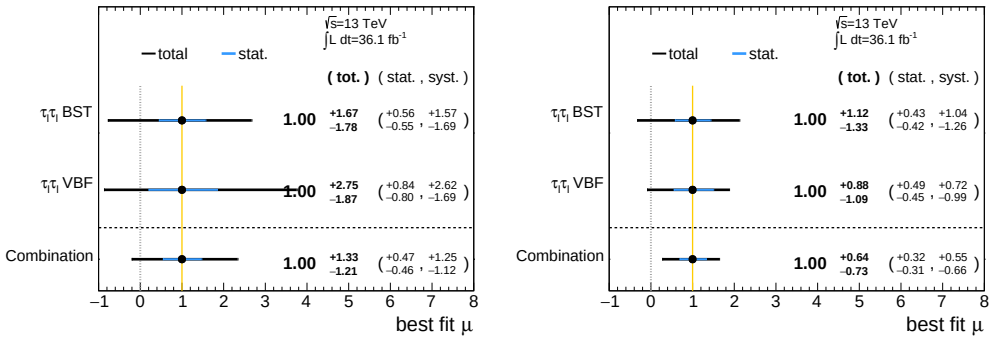
Figure 11.3.: Distributions of the BDT output in the different signal regions with the binning used in the fit of the multivariate analysis. For the data shown in the plot Asimov data corresponding to the 2015 and 2016 dataset with 36.1 fb^{-1} are used.



(a) $Z \to \ell\ell$ boosted **(b)** $Z \to \ell\ell$ VBF

(c) Top boosted **(d)** Top VBF

Figure 11.4.: Distributions of the BDT output in the different control regions with the binning used in the fit of the multivariate analysis. For the data shown in the plot Asimov data corresponding to the 2015 and 2016 dataset with 36.1 fb^{-1} are used.



(a) CBA **(b)** MVA

Figure 11.5.: Uncertainties on the signal strength measured in the cut-based analysis (a) and multivariate analysis (b) with Asimov data corresponding to the 2015 and 2016 dataset with 36.1 fb^{-1}.

Table 11.4.: Uncertainty on the signal strength measurement in the cut-based (CBA) and multivariate (MVA) analysis with Asimov data corresponding to the 2015 and 2016 dataset with 36.1 fb^{-1}.

Source of uncertainty	Impact on $\Delta\mu$ error	
	CBA	MVA
Total	±1.27	±0.68
Data statistics	±0.46	±0.32
Systematic uncertainties	±1.18	±0.60
Normalization uncertainties	±0.62	±0.25
Jets and $E_{\text{T}}^{\text{miss}}$	±0.92	±0.45
b-jets	±0.21	±0.03
Light leptons	±0.12	±0.18
Pileup reweighting	±0.11	±0.04
Fake estimation	±0.17	±0.04
Luminosity	±0.04	±0.04
Theory unc. on signal	±0.21	±0.09
Theory unc. on $Z \to \tau\tau$	±0.19	±0.07
Simulation statistics	±0.59	±0.32

12. Conclusion and Outlook

In this thesis the measurement of the analysis of the $H \to \tau^+\tau^- \to \ell^+\ell^- 4\nu$ decay process is presented. The full 2015 and 2016 dataset corresponding to $36.1\,\text{fb}^{-1}$ recorded with the ATLAS detector in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$ is used. After introducing the event selection of the cut-based analysis a method to use boosted decision trees (BDTs) in the analysis in order to increase the sensitivity on the signal is developed.

First, the event selection of the cut-based analysis is modified to allow for more inclusive signal regions. The idea is, that the separation between signal and background events is not achieved by applying requirements on the events but to use a multivariate approach in the form of BDTs. The categorization into the VBF and boosted category, which are used to select the VBF and ggF production modes of the Higgs boson, is preserved. However, the categories are further split into a same flavour (SF, ee and $\mu\mu$) and different flavour (DF, $e\mu$ and μe) region based on the flavour of the leptons in the final state, which allows a better training of the BDTs because those regions have different background compositions.

Boosted decision trees need some data where the correct classification in signal or background is known to be trained on. This information is only provided by simulated events. But the simulated events are also required in the statistical analysis, where a prediction of expected signal and background yields is needed. The same set of events cannot be used for the training and the fit, because the BDTs could have used statistical fluctuations to increase the separation power between signal and background. Therefore, two statistically independent sets need too be used for training and fitting, which are called *training* and *test* set, respectively. The performance of BDTs depends on the *hyperparameters*, which control the behavior of the training. To estimate the performance of BDTs with different hyperparameters a third statistically independent set is needed, which is called the *validation* set. If the simulated events are just split into these three sets, both the BDT training and statistical analysis suffer from low statistics. The k-fold cross-validation approach is used to increase the training statistics by splitting the simulated events into k independent slices. Now, k BDTs are trained, each using $k-2$ slices for training, one slice for validation, and one slice for testing. The slices are assigned in such a ways, that every slice is used $k-2$ times for training, and once for validation and testing. The validation and test steps are performed on the whole set of simulated events by combining all k BDTs. For this thesis the value $k = 10$ is chosen.

The following hyperparameters are optimized: boosting algorithm, number of trees used in the boosting, maximum depth of one decision tree, minimum number of events in the final nodes of the decision tree, and learning rate of the boosting

algorithm. The best performing BDT is selected based on the sensitivity which is calculated using the statistical analysis without consideration of systematic uncertainties. Additionally, a threshold on the KS-test probability is set to discard BDTs with too much overtraining. In all regions the best performing BDTs use the gradient boost algorithm with $250 - 500$ boosting steps, a small depth, and a slow learning rate.
In a second optimization step, the number of observables which are used as input variables for the BDTs is optimized in an iterative way. In every step the influence of each used observable on the splitting power of the BDT is calculated and the least performing observable is discarded. Based on the drop in expected significance the final observable count is determined. In the end 9 and 8 variables are used for the BDTs in the VBF SF and VBF DF region, whereas the BDTs in the boosted SF and boosted DF region use 4 and 9 variables, respectively.
The expected sensitivity of the multivariate analysis is calculated in a likelihood fit and compared to the expected sensitivity of the cut-based analysis. To improve the sensitivity the binning of the BDT distributions is optimized. An expected sensitivity of 0.83σ is reached in the cut-based analysis, whereas the multivariate analysis resulted in an expected sensitivity incerease by 63 % to 1.35σ. The attempt to increase the sensitivity with multivariate techniques was therefore successful. Additionally, the expected uncertainty on the measured signal strength is also reduced by a factor of two from 1 ± 1.27 in the cut-based analysis to 1 ± 0.68 in the multivariate analysis.
Of course, there is still room for further studies and improvements. The categorization of the ggF and VBF production modes is achieved by requirements on different observables. Here also multivariate techniques in form of BDTs could be applied. A first BDT can be used to separate signal and background events and the splitting between ggF and VBF production mode is accomplished by a second BDT. In general other multivariate techniques like neutral networks could be used instead of BDTs. It is known, that neural networks achieve a similar or even better separation power [127], which could be used to increase the sensitivity even more. Also, the k-fold cross-validation approach could be improved. Currently data events have also to be split into k sets. This can be avoided when using only half the simulated events for the k-fold cross validation. After fixing the parameters and input variables a 2-fold cross-evaluation scheme can be applied on the two sets of simulated events. In this way, data events only need to be split in two sets.

A. Additional Material for Chapter 9

A.1. Observables used in the optimization

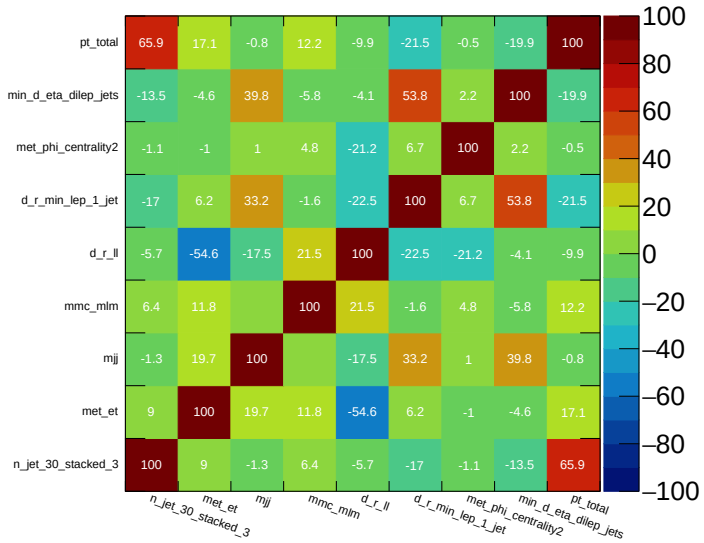
Table A.1.: List of all observables considered in the optimization of the multivariate analysis.

Variable	Description
m_{MMC}	output of the missing mass calculator, see Section 6.1.2
$m_{\ell\ell}$	invariant mass of the dilepton system
m_{jj}	invariant mass of the system of the first two leading jets
$m_{\tau\tau,\text{j}_1}$	sum of the mass of the di-τ system and the leading jet
$p_{\text{T}}^{\ell_1}$	transverse momentum of the leading lepton
η_{ℓ_1}	η of the leading lepton
$m_{\text{T}}^{\ell_1}$	transverse mass of the leading lepton with the transverse energy
$p_{\text{T}}^{\ell_2}$	transverse momentum of the subleading lepton
η_{ℓ_2}	η of the subleading lepton
$m_{\text{T}}^{\ell_2}$	transverse mass of the leading lepton with the transverse energy
n_{jets}	number of jets above 30 GeV, bins for $n_{\text{jets}} \geq 3$ are merged
$p_{\text{T}}^{\text{j}_1}$	transverse momentum of the leading jet
η_{j_1}	η of the leading jet
$p_{\text{T}}^{\text{j}_2}$	transverse momentum of the subleading jet
η_{j_2}	η of the subleading jet
$p_{\text{T}}^{\text{j}_3}$	transverse momentum of the third jet
η_{j_3}	η of the third jet
$E_{\text{T}}^{\text{miss}}$	missing transverse energy
$\sum \lvert p_{\text{T}} \rvert$	scalar sum of $p_{\text{T}}^{\ell_1}$, $p_{\text{T}}^{\ell_2}$, $p_{\text{T}}^{\text{jet}_1}$, $p_{\text{T}}^{\text{jet}_2}$, and $E_{\text{T}}^{\text{miss}}$
$p_{\text{T}}^{\text{total}}$	vectorial sum of $p_{\text{T}}^{\ell_1}$, $p_{\text{T}}^{\ell_2}$, $p_{\text{T}}^{\text{jet}_1}$, $p_{\text{T}}^{\text{jet}_2}$, and $E_{\text{T}}^{\text{miss}}$
$E_{\text{T,x}}^{\text{miss,HPTO}}$	x component of the high-p_{T} object based missing transverse energy
$E_{\text{T,y}}^{\text{miss,HPTO}}$	y component of the high-p_{T} object based missing transverse energy
$E_{\text{T},\phi}^{\text{miss,HPTO}}$	ϕ component of the high-p_{T} object based missing transverse energy
x_1^{HPTO}	Eq. (6.2) for high-p_{T} objects
x_2^{HPTO}	Eq. (6.2) for high-p_{T} objects
$p_{\text{T}}^{\tau\tau}$	transverse momentum of the di-τ system
$p_{\text{T}}^{\ell\ell}$	transverse momentum of the dilepton system
$p_{\text{T}}^{\tau\tau,\text{mcoll}}$	transverse momentum of the di-τ system in the collinear approximation

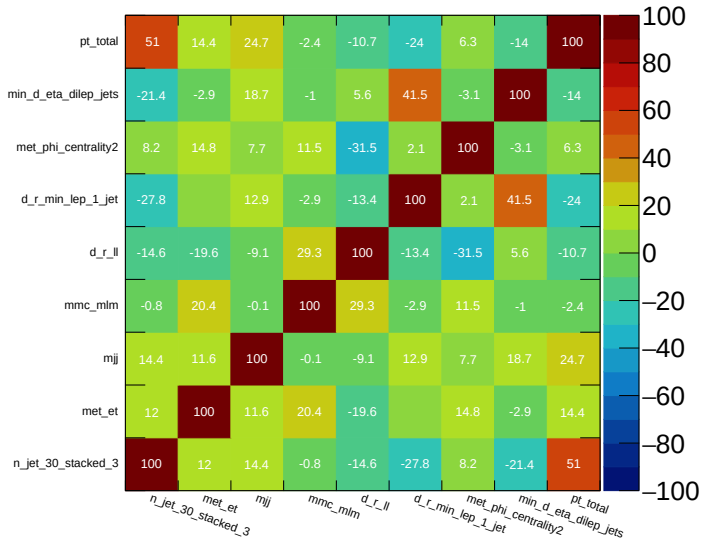
Table A.2.: List of all observables considered in the optimization of the multivariate analysis (continued).

Variable	Description
$p_{\mathrm{T}}(\ell_1 + E_{\mathrm{T}}^{\mathrm{miss}})$	transverse momentum of the system of the leading lepton and $E_{\mathrm{T}}^{\mathrm{miss}}$
$p_{\mathrm{T}}(\ell_2 + E_{\mathrm{T}}^{\mathrm{miss}})$	transverse momentum of the system of the subleading lepton and $E_{\mathrm{T}}^{\mathrm{miss}}$
x_1	momentum fraction of leading lepton w.r.t. to τ-lepton, see Eq. (6.2)
x_2	momentum fraction of leading lepton w.r.t. to τ-lepton, see Eq. (6.2)
$\lvert\Delta\phi_{\ell\ell}\rvert$	absolute value of the distance in ϕ between the leading leptons
$\Delta\phi(\ell_1, E_{\mathrm{T}}^{\mathrm{miss}})$	distance in ϕ between the leading lepton and $E_{\mathrm{T}}^{\mathrm{miss}}$
$\Delta\phi(\ell_2, E_{\mathrm{T}}^{\mathrm{miss}})$	distance in ϕ between the subleading lepton and $E_{\mathrm{T}}^{\mathrm{miss}}$
$\Delta\eta_{\ell\ell}$	distance in η between the leading leptons
$\Delta\eta_{\mathrm{jj}}$	distance in η between the leading jets
$\min\Delta R(\ell_1, \mathrm{jets})$	minimal ΔR distance between the leading lepton and all jets
$\min\Delta R(\ell_2, \mathrm{jets})$	minimal ΔR distance between the subleading lepton and all jets
$\min\Delta R(\ell\ell, \mathrm{jets})$	minimal ΔR distance between the dilepton system and all jets
$C_\eta\ell_1$	η centrality between the two leading jets and the leading lepton [134]
$C_\eta\ell_2$	η centrality between the two leading jets and the subleading lepton
$C_\eta\ell_1 \cdot C_\eta\ell_2$	product of the η centralities above
$C_\eta\mathrm{j}_3$	η centrality between the two leading jets and the third jet
$E_{\mathrm{T}}^{\mathrm{miss}}\phi$ centrality	see Eq. (9.5)
$\eta_{\mathrm{j}_1} \cdot \eta_{\mathrm{j}_2}$	product of η of leading and subleading jet
$\left\lvert\eta_{\mathrm{j}_1} - \eta_{\mathrm{j}_2}\right\rvert$	absolute distance in η between the two leading jets
$\Delta\eta(\mathrm{j}_1\mathrm{j}_2, \mathrm{j}_3)$	distance in η between the system of the two leading jets and the third jet
$E_{\mathrm{T}}^{\mathrm{miss}}/p_{\mathrm{T}}^{\ell_1}$	ratio of $E_{\mathrm{T}}^{\mathrm{miss}}$ and $p_{\mathrm{T}}^{\ell_1}$
$E_{\mathrm{T}}^{\mathrm{miss}}/p_{\mathrm{T}}^{\ell_2}$	ratio of $E_{\mathrm{T}}^{\mathrm{miss}}$ and $p_{\mathrm{T}}^{\ell_2}$
$p_{\mathrm{T}}^{\ell_1}/p_{\mathrm{T}}^{\ell_2}$	ratio of $p_{\mathrm{T}}^{\ell_1}$ and $p_{\mathrm{T}}^{\ell_2}$
Sphericity	see Eq. (9.9)
$\eta^*_{\mathrm{j}_3}$	the Zeppenfeld variable for the third jet [149]
$\eta^*_{\ell\ell}$	the Zeppenfeld variable for the dilepton system

A.2. Correlations of input variables

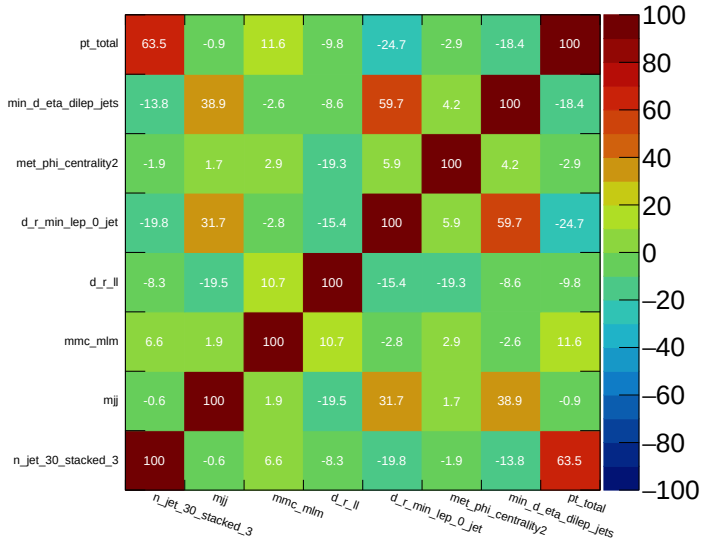


(a) Signal.

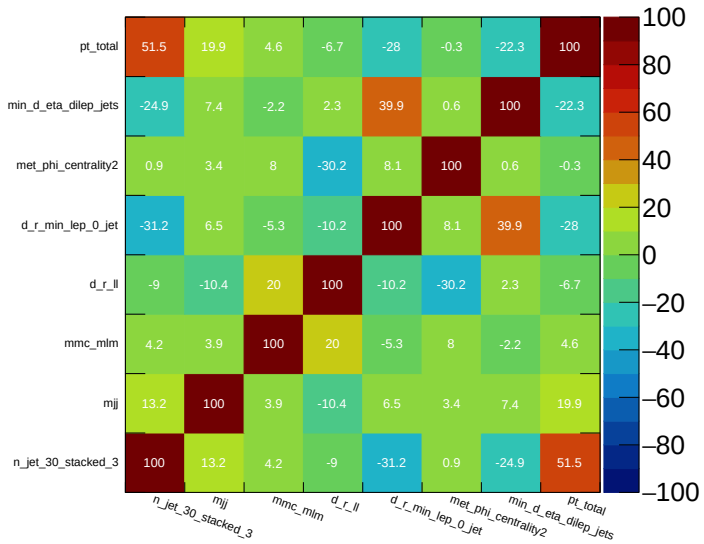


(b) Background

Figure A.1.: Correlations of the input variables for the BDTs in the VBF SF category for signal and background events.

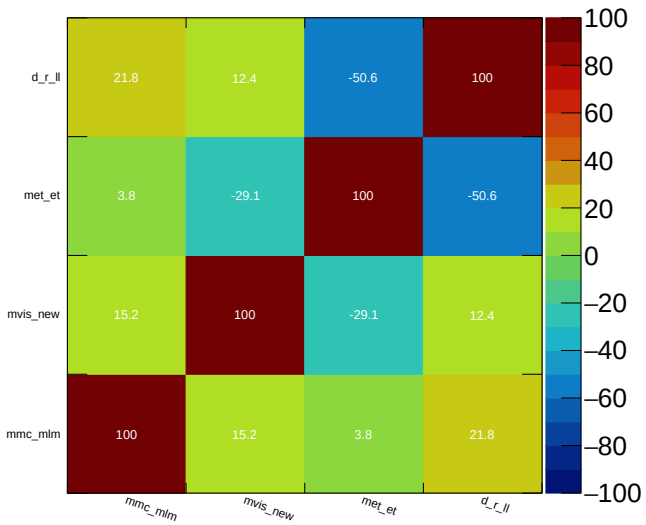


(a) Signal.

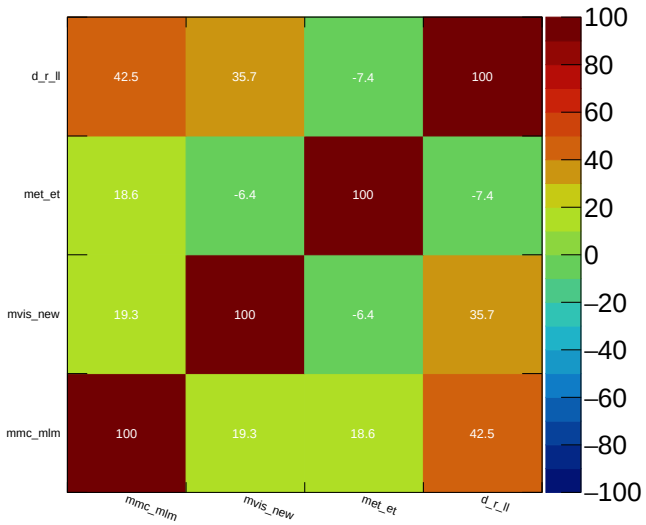


(b) Background

Figure A.2.: Correlations of the input variables for the BDTs in the VBF DF category for signal and background events.

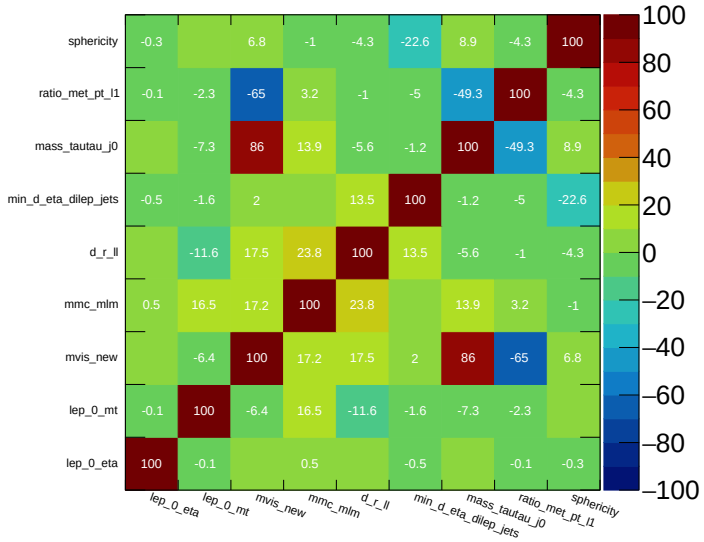


(a) Signal.

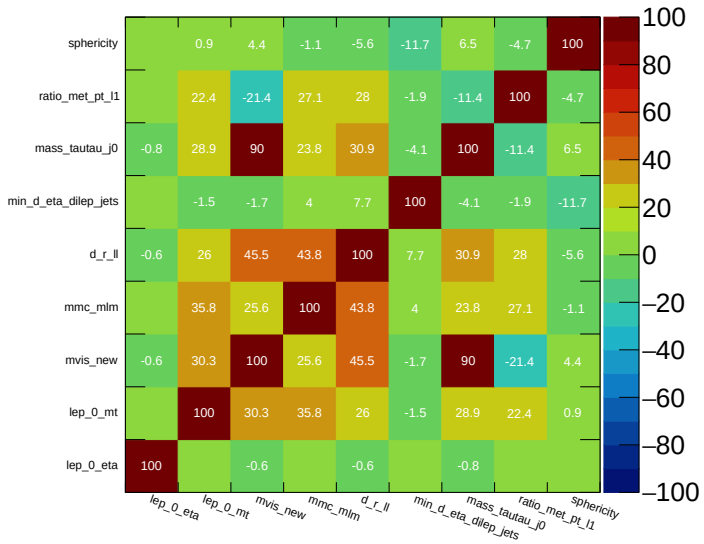


(b) Background

Figure A.3.: Correlations of the input variables for the BDTs in the boosted SF category for signal and background events.



(a) Signal.



(b) Background

Figure A.4.: Correlations of the input variables for the BDTs in the boosted DF category for signal and background events.

Bibliography

[1] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. **B716** (2012) 1–29, arXiv:1207.7214 [hep-ex].

[2] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. **B716** (2012) 30–61, arXiv:1207.7235 [hep-ex].

[3] ATLAS and CMS Collaboration, *Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments*, Phys. Rev. Lett. **114** (2015) 191803, arXiv:1503.07589 [hep-ex].

[4] UA1 Collaboration, *Experimental observation of lepton pairs of invariant mass around* $95\,\mathrm{GeV}/_{\mathrm{c}}^{2}$ *at the CERN SPS collider*, Physics Letters B **126** (1983) 398–410.

[5] UA1 Collaboration, *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.

[6] UA2 Collaboration, *Evidence for* $Z^0 \to e^+e^-$ *at the CERN pp collider*, Physics Letters B **129** (1983) 130–140.

[7] UA2 Collaboration, *Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN pp collider*, Physics Letters B **122** (1983) 476–485.

[8] Particle Data Group Collaboration, C. Patrignani et al., *Review of Particle Physics*, Chin. Phys. **C40** (2016) 100001.

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

[10] P. W. Higgs, *Spontaneous Symmetry Breakdown without Massless Bosons*, Phys. Rev. **145** (1966) 1156–1163.

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

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

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

[14] T. W. B. Kibble, *Symmetry Breaking in Non-Abelian Gauge Theories*, Phys. Rev. **155** (1967) 1554–1561.

[15] ATLAS Collaboration, *Evidence for the Higgs-boson Yukawa coupling to tau leptons with the ATLAS detector*, JHEP **04** (2015) 117, `arXiv:1501.04943 [hep-ex]`.

[16] CBF Collaboration, *Observation of Top Quark Production in $\overline{p}p$ Collisions with the Collider Detector at Fermilab*, Phys. Rev. Lett. **74** (1995) 2626–2631.

[17] D0 Collaboration, *Observation of the top quark*, Phys. Rev. Lett. **74** (1995) 2632–2637, `arXiv:hep-ex/9503003 [hep-ex]`.

[18] D. Griffiths, *Introduction to Elementary Particles*. WILEY-VCH Verlag GmbH & Co. KGaA, second ed., 2008. ISBN: 978-3-527-40601-2.

[19] F. Halzen and A. D. Martin, *Quarks and Leptons: An Introductory Course in Modern Particle Physics*. Wiley, 1984. ISBN: 978-0471887416.

[20] M. E. Peskin and D. V. Schroeder, *An Introduction To Quantum Field Theory*. Westview Press, 1995. ISBN: 0201503972.

[21] G. V. Chibisov, *Astrophysical upper limits on the photon rest mass*, Soviet Physics Uspekhi **19** (1976) 624.

[22] P. Schmüser, *Feynman-Graphen und Eichtheorien für Experimentalphysiker*. Springer, Berlin, Heidelberg, 1995. ISBN: 978-3-642-57766-6.

[23] D. J. Gross and F. Wilczek, *Ultraviolet Behavior of Non-Abelian Gauge Theories*, Phys. Rev. Lett. **30** (1973) 1343–1346.

[24] H. D. Politzer, *Reliable Perturbative Results for Strong Interactions?*, Phys. Rev. Lett. **30** (1973) 1346–1349.

[25] T. D. Lee and C. N. Yang, *Question of Parity Conservation in Weak Interactions*, Phys. Rev. **104** (1956) 254–258.

[26] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson, *Experimental Test of Parity Conservation in Beta Decay*, Phys. Rev. **105** (1957) 1413–1415.

[27] S. L. Glashow, *Partial-symmetries of weak interactions*, Nuclear Physics **22** (1961) 579 – 588.

[28] A. Salam and J. C. Ward, *Weak and electromagnetic interactions*, Il Nuovo Cimento (1955-1965) **11** (1959) 568–577.

[29] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264–1266.

[30] K. Nishijima, *Charge of Independence Theory of V Particles*, Progress of Theoretical Physics **13** (1955) 285–304.

[31] M. Gell-Mann, *The interpretation of the new particles as displaced charge multiplets*, Il Nuovo Cimento (1955-1965) **4** (1956) 848–866.

[32] Z. Maki, M. Nakagawa, and S. Sakata, *Remarks on the Unified Model of Elementary Particles*, Progress of Theoretical Physics **28** (1962) 870–880.

[33] B. Pontecorvo, *Neutrino Experiments and the Problem of Conservation of Leptonic Charge*, Sov. Phys. JETP **26** (1968) 984–988, [Zh. Eksp. Teor. Fiz.53,1717(1967)].

[34] Super-Kamiokande Collaboration, Y. Fukuda et al., *Evidence for oscillation of atmospheric neutrinos*, Phys. Rev. Lett. **81** (1998) 1562–1567, `arXiv:hep-ex/9807003 [hep-ex]`.

[35] SNO Collaboration, Q. R. Ahmad et al., *Measurement of the Rate of $\nu_e + d \to p + p + e^-$ Interactions Produced by ^8B Solar Neutrinos at the Sudbury Neutrino Observatory*, Phys. Rev. Lett. **87** (2001) 071301.

[36] SNO Collaboration, Q. R. Ahmad et al., *Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory*, Phys. Rev. Lett. **89** (2002) 011301.

[37] K2K Collaboration, *Measurement of Neutrino Oscillation by the K2K Experiment*, Phys. Rev. **D74** (2006) 072003, `arXiv:hep-ex/0606032 [hep-ex]`.

[38] M. Kobayashi and T. Maskawa, *CP-Violation in the Renormalizable Theory of Weak Interaction*, Progress of Theoretical Physics **49** (1973) 652–657.

[39] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10** (1963) 531–533.

[40] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, *Evidence for the 2π Decay of the K_2^0 Meson*, Phys. Rev. Lett. **13** (1964) 138–140.

[41] M. Schumacher, *Suche nach neutralen Higgs-Bosonen mit dem OPAL-Detektor am LEP2*. PhD thesis, Rheinische Friedrich-Wilhelms-Universität Bonn, 1999. [BONN-IR-99-19].

[42] Y. Nambu, *Quasi-Particles and Gauge Invariance in the Theory of Superconductivity*, Phys. Rev. **117** (1960) 648–663.

[43] J. Goldstone, *Field theories with « Superconductor » solutions*, Il Nuovo Cimento (1955-1965) **19** (1961) 154–164.

[44] ATLAS Collaboration, *ATLAS TWiki*, `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/StandardModelPublicResults`. [Online; accessed 15-October-2017].

[45] S. D. Drell and T.-M. Yan, *Partons and their applications at high energies*, Annals of Physics **66** (1971) 578 – 623.

[46] LHC Higgs Cross Section Working Group Collaboration, D. de Florian et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, `arXiv:1610.07922 [hep-ph]`.

[47] LHC Higgs Cross Section Working Group Collaboration, J. R. Andersen et al., *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*, `arXiv:1307.1347 [hep-ph]`.

[48] ATLAS and CMS Collaboration, *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*, JHEP **08** (2016) 045, `arXiv:1606.02266 [hep-ex]`.

[49] LHC Higgs Cross Section Working Group Collaboration, S. Dittmaier et al., *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*, `arXiv:1101.0593 [hep-ph]`.

[50] ATLAS Collaboration, *Measurement of the Higgs boson mass from the* $H \to \gamma\gamma$ *and* $H \to ZZ^* \to 4\ell$ *channels with the ATLAS detector using 25* fb^{-1} *of pp collision data*, Phys. Rev. **D90** (2014) 052004, `arXiv:1406.3827 [hep-ex]`.

[51] CMS Collaboration, CMS Collaboration, *Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV*, Eur. Phys. J. **C75** (2015) 212, `arXiv:1412.8662 [hep-ex]`.

[52] CMS Collaboration, *Limits on the Higgs boson lifetime and width from its decay to four charged leptons*, Phys. Rev. **D92** (2015) 072010, `arXiv:1507.06656 [hep-ex]`.

[53] CMS Collaboration, CMS Collaboration, *Constraints on the Higgs boson width from off-shell production and decay to Z-boson pairs*, Phys. Lett. **B736** (2014) 64–85, `arXiv:1405.3455 [hep-ex]`.

[54] CMS Collaboration, *Search for Higgs boson off-shell production in proton-proton collisions at 7 and 8 TeV and derivation of constraints on its total decay width*, JHEP **09** (2016) 051, `arXiv:1605.02329 [hep-ex]`.

[55] ATLAS Collaboration, *Constraints on the off-shell Higgs boson signal strength in the high-mass ZZ and WW final states with the ATLAS detector*, Eur. Phys. J. **C75** (2015) 335, `arXiv:1503.01060 [hep-ex]`.

[56] ATLAS Collaboration, ATLAS Collaboration, *Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector*, Eur. Phys. J. **C75** (2015) 476, arXiv:1506.05669 [hep-ex], [Erratum: Eur. Phys. J.C76,no.3,152(2016)].

[57] CMS Collaboration, *Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV*, Phys. Rev. **D92** (2015) 012004, arXiv:1411.3441 [hep-ex].

[58] ATLAS Collaboration, *Test of CP Invariance in vector-boson fusion production of the Higgs boson using the Optimal Observable method in the ditau decay channel with the ATLAS detector*, Eur. Phys. J. **C76** (2016) 658, arXiv:1602.04516 [hep-ex].

[59] ATLAS Collaboration, *Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector*, Eur. Phys. J. **C75** (2015) 476, arXiv:1506.05669 [hep-ex], [Erratum: Eur. Phys. J.C76,no.3,152(2016)].

[60] CMS Collaboration, *Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV*, Phys. Rev. **D92** (2015) 012004, arXiv:1411.3441 [hep-ex].

[61] ATLAS Collaboration, *Combined measurements of Higgs boson production and decay in the $H \to ZZ^* \to 4\ell$ and $H \to \gamma\gamma$ channels using $\sqrt{s} = 13$ TeV proton–proton collision data collected with the ATLAS experiment*, ATLAS-CONF-2017-047, 2017, https://cds.cern.ch/record/2273854.

[62] ATLAS Collaboration, *Measurement of the Higgs boson mass in the $H \to ZZ^* \to 4\ell$ and $H \to \gamma\gamma$ channels with $\sqrt{s} = 13$ TeV pp collisions using the ATLAS detector*, ATLAS-CONF-2017-046, 2017, https://cds.cern.ch/record/2273853.

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

[64] The ATLAS Collaboration et al., *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[65] The CMS Collaboration et al., *The CMS experiment at the CERN LHC*, JINST **3** (2008) S08004.

[66] The LHCb Collaboration et al., *The LHCb Detector at the LHC*, JINST **3** (2008) S08005.

[67] The ALICE Collaboration et al., *The ALICE experiment at the LHC*, JINST **3** (2008) S08002.

[68] E. Mobs, *The CERN accelerator complex. Complexe des accélérateurs du CERN*, https://cds.cern.ch/record/2197559, General Photo.

[69] ATLAS Collaboration, *ATLAS TWiki*, https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults. [Online; accessed 02-October-2017].

[70] ATLAS Collaboration, *ATLAS TWiki*, https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2. [Online; accessed 02-October-2017].

[71] ATLAS Collaboration, *ATLAS Insertable B-Layer Technical Design Report*, Atlas-tdr-19, 2010, https://cds.cern.ch/record/1291633.

[72] ATLAS Collaboration, *ATLAS TWiki*, https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ApprovedPlotsDAQ. [Online; accessed 02-October-2017].

[73] ATLAS Collaboration, *ATLAS TWiki*, https://twiki.cern.ch/twiki/bin/view/AtlasPublic/TriggerOperationPublicResults. [Online; accessed 02-October-2017].

[74] ATLAS Collaboration, *ATLAS TWiki*, https://twiki.cern.ch/twiki/bin/viewauth/Atlas/LuminosityForPhysics. [Online; accessed 17-October-2017].

[75] S. Alioli, P. Nason, C. Oleari, and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010) 043, arXiv:1002.2581 [hep-ph].

[76] T. Sjöstrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, arXiv:0710.3820 [hep-ph].

[77] K. Hamilton, P. Nason, E. Re, and G. Zanderighi, *NNLOPS simulation of Higgs boson production*, JHEP **10** (2013) 222, arXiv:1309.0017 [hep-ph].

[78] J. Pumplin et al., *New generation of parton distributions with uncertainties from global QCD analysis*, JHEP **07** (2002) 012, arXiv:hep-ph/0201195.

[79] NNPDF Collaboration, R. D. Ball et al., *Parton distributions for the LHC Run II*, JHEP **04** (2015) 040, arXiv:1410.8849 [hep-ph].

[80] T. Gleisberg, S. Höche, F. Krauss, M. Schönherr, S. Schumann, et al., *Event generation with SHERPA 1.1*, JHEP **02** (2009) 007, arXiv:0811.4622 [hep-ph].

[81] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, JHEP **12** (2008) 039, arXiv:0808.3674 [hep-ph].

[82] F. Cascioli, P. Maierhofer, and S. Pozzorini, *Scattering Amplitudes with Open Loops*, Phys. Rev. Lett. **108** (2012) 111601, arXiv:1111.5206 [hep-ph].

[83] S. Schumann and F. Krauss, *A Parton shower algorithm based on Catani-Seymour dipole factorisation*, JHEP **03** (2008) 038, arXiv:0709.1027 [hep-ph].

[84] S. Hoeche, F. Krauss, M. Schonherr, and F. Siegert, *QCD matrix elements + parton showers: The NLO case*, JHEP **04** (2013) 027, arXiv:1207.5030 [hep-ph].

[85] T. Sjöstrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **05** (2006) 026, arXiv:hep-ph/0603175.

[86] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70** (2010) 823, arXiv:1005.4568 [hep-ex].

[87] GEANT4 Collaboration, S. Agostinelli et al., *GEANT4: A simulation toolkit*, Nucl. Instrum. Meth. A **506** (2003) 250.

[88] S. Catani, L. Cieri, G. Ferrera, D. de Florian, and M. Grazzini, *Vector boson production at hadron colliders: a fully exclusive QCD calculation at NNLO*, Phys. Rev. Lett. **103** (2009) 082001, arXiv:0903.2120 [hep-ph].

[89] S. Catani and M. Grazzini, *Next-to-Next-to-Leading-Order Subtraction Formalism in Hadron Collisions and its Application to Higgs-Boson Production at the Large Hadron Collider*, Phys. Rev. Lett. **98** (2007) 222002.

[90] T. Binoth, M. Ciccolini, N. Kauer, and M. Kramer, *Gluon-induced W-boson pair production at the LHC*, JHEP **12** (2006) 046, arXiv:hep-ph/0611170 [hep-ph].

[91] J. M. Campbell, R. K. Ellis, and C. Williams, *Vector boson pair production at the LHC*, JHEP **07** (2011) 018, arXiv:1105.0020 [hep-ph].

[92] N. Kidonakis, *NNLL resummation for s-channel single top quark production*, Phys. Rev. **D81** (2010) 054028, arXiv:1001.5034 [hep-ph].

[93] N. Kidonakis, *Two-loop soft anomalous dimensions for single top quark associated production with a W- or H-*, Phys. Rev. **D82** (2010) 054018, arXiv:1005.4451 [hep-ph].

[94] N. Kidonakis, *Two-loop soft anomalous dimensions for single top quark associated production with a W- or H-*, Phys. Rev. **D82** (2010) 054018, arXiv:1005.4451 [hep-ph].

[95] M. Cacciari, M. Czakon, M. Mangano, A. Mitov, and P. Nason, *Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation*, Phys. Lett. **B710** (2012) 612–622, arXiv:1111.5869 [hep-ph].

[96] P. Bärnreuther, M. Czakon, and A. Mitov, *Percent Level Precision Physics at the Tevatron: First Genuine NNLO QCD Corrections to $q\bar{q} \to t\bar{t} + X$*, Phys. Rev. Lett. **109** (2012) 132001, arXiv:1204.5201 [hep-ph].

[97] M. Czakon and A. Mitov, *NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels*, JHEP **12** (2012) 054, arXiv:1207.0236 [hep-ph].

[98] M. Czakon and A. Mitov, *NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction*, JHEP **01** (2013) 080, arXiv:1210.6832 [hep-ph].

[99] M. Czakon, P. Fiedler, and A. Mitov, *Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(a_S^4)$*, Phys. Rev. Lett. **110** (2013) 252004, arXiv:1303.6254 [hep-ph].

[100] T. Cornelissen, M. Elsing, S. Fleischmann, W. Liebig, E. Moyse, and A. Salzburger, *Concepts, Design and Implementation of the ATLAS New Tracking (NEWT)*, Tech. Rep. ATL-SOFT-PUB-2007-007. ATL-COM-SOFT-2007-002, CERN, Geneva, Mar, 2007. https://cds.cern.ch/record/1020106.

[101] ATLAS Collaboration, *Early Inner Detector Tracking Performance in the 2015 Data at $\sqrt{s} = 13$ TeV*, ATL-PHYS-PUB-2015-051, 2015, https://cds.cern.ch/record/2110140.

[102] ATLAS Collaboration, *Vertex Reconstruction Performance of the ATLAS Detector at $\sqrt{s} = 13$ TeV*, ATL-PHYS-PUB-2015-026, 2015, https://cds.cern.ch/record/2037717.

[103] ATLAS Collaboration, *The Optimization of ATLAS Track Reconstruction in Dense Environments*, ATL-PHYS-PUB-2015-006, 2015, https://cds.cern.ch/record/2002609.

[104] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton–proton collision data*, ATLAS-CONF-2016-024, 2016, https://cds.cern.ch/record/2157687.

[105] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **76** (2016) 292, arXiv:1603.05598 [hep-ex].

[106] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04** (2008) 063, arXiv:0802.1189 [hep-ph].

[107] M. Cacciari and G. P. Salam, *Dispelling the N^3 myth for the k_t jet-finder*, Phys. Lett. B **641** (2006) 57, arXiv:hep-ph/0512210.

[108] ATLAS Collaboration, *Jet energy scale measurements and their systematic uncertainties in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, arXiv:1703.09665 [hep-ex].

[109] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector*, Eur. Phys. J. C **76** (2016) 581, arXiv:1510.03823 [hep-ex].

[110] ATLAS Collaboration, *Forward Jet Vertex Tagging: A new technique for the identification and rejection of forward pileup jets*, ATL-PHYS-PUB-2015-034, 2015, https://cds.cern.ch/record/2042098.

[111] ATLAS Collaboration, *Performance of b-jet identification in the ATLAS experiment*, JINST **11** (2016) P04008, arXiv:1512.01094 [hep-ex].

[112] ATLAS Collaboration, *Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run*, ATL-PHYS-PUB-2016-012, 2016, https://cds.cern.ch/record/2160731.

[113] ATLAS Collaboration, *Identification and energy calibration of hadronically decaying tau leptons with the ATLAS experiment in pp collisions at $\sqrt{s} = 8$ TeV*, Eur. Phys. J. C **75** (2015) 303, arXiv:1412.7086 [hep-ex].

[114] ATLAS Collaboration, *Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC*, https://cds.cern.ch/record/2064383.

[115] ATLAS Collaboration, *Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at $\sqrt{s} = 13$ TeV*, ATLAS-CONF-2017-029, 2017, https://cds.cern.ch/record/2261772.

[116] J. D. Bjorken and E. A. Paschos, *Inelastic Electron-Proton and γ-Proton Scattering and the Structure of the Nucleon*, Phys. Rev. **185** (1969) 1975–1982.

[117] ATLAS Collaboration, *Expected performance of missing transverse momentum reconstruction for the ATLAS detector at $\sqrt{s} = 13$ TeV*, ATL-PHYS-PUB-2015-023, 2015, https://cds.cern.ch/record/2037700.

[118] ATLAS Collaboration, *Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run-1 data*, arXiv:1606.01813 [hep-ex].

[119] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector in the first proton–proton collisions at $\sqrt{s} = 13$ TeV*, ATL-PHYS-PUB-2015-027, 2015, https://cds.cern.ch/record/2037904.

[120] R. Ellis, I. Hinchliffe, M. Soldate, and J. V. D. Bij, *Higgs decay to $\tau^+\tau^-$ A possible signature of intermediate mass Higgs bosons at high energy hadron colliders*, Nuclear Physics B **297** (1988) 221 – 243.

[121] ATLAS Collaboration, G. Aad et al., *Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics*, arXiv:0901.0512 [hep-ex].

[122] A. Elagin, P. Murat, A. Pranko, and A. Safonov, *A New Mass Reconstruction Technique for Resonances Decaying to di-tau*, Nucl. Instrum. Meth. **A654** (2011) 481–489, arXiv:1012.4686 [hep-ex].

[123] A. Andreazza et al., *Measurement of the $H \to \tau^+\tau^-$ cross-section in* 13 TeV *Collisions with the ATLAS Detector*, Tech. Rep. ATL-COM-PHYS-2017-446, CERN, Geneva, Apr, 2017. https://cds.cern.ch/record/2261605.

[124] ATLAS Collaboration, M. Aaboud et al., *Performance of the ATLAS Trigger System in 2015*, Eur. Phys. J. **C77** (2017) 317, arXiv:1611.09661 [hep-ex].

[125] ATLAS Collaboration, *Trigger Menu in 2016*, Tech. Rep. ATL-DAQ-PUB-2017-001, CERN, Geneva, Jan, 2017. https://cds.cern.ch/record/2242069.

[126] R. Brun and F. Rademakers, *ROOT - An object oriented data analysis framework*, Proceedings AIHENP'96 Workshop, Lausanne, Sep. 1996, Nucl. Inst. & Meth. in Phys. Res. A **389** (1997) 81 – 86, See also http://root.cern.ch/.

[127] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss, *TMVA: Toolkit for Multivariate Data Analysis*, PoS **ACAT** (2007) 040, arXiv:physics/0703039.

[128] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data mining, Inference and Prediction*. Springer, 2 ed., 2009. ISBN: 978-0-387-84858-7.

[129] 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.

[130] G. Cowan and E. Gross, *Discovery significance with statistical uncertainty in the background estimate*, https://www.pp.rhul.ac.uk/~cowan/stat/notes/SigCalcNote.pdf, May, 2008. [Online; accessed 13-October-2017].

[131] A. Kolmogorov, *Sulla Determinazione Empirica di una Legge di Distribuzione*, Giornale dell'Istituto Italiano degli Attuari **4** (1933) 83 – 91.

[132] N. Smirnov, *Table for Estimating the Goodness of Fit of Empirical Distributions*, The Annals of Mathematical Statistics **19** (1948) 279 – 281.

[133] L. Breiman, J. Friedman, C. Stone, and R. Olshen, *Classification and Regression Trees*. The Wadsworth and Brooks-Cole statistics-probability series. Taylor & Francis, 1984. ISBN: 9780412048418.

[134] C. Schillo, *Search for the Standard Model Higgs Boson and Test of CP Invariance in the Vector-Boson Fusion Production of the Higgs Boson in the Fully Leptonic $H \to \tau^+\tau^- \to \ell^+\ell^- 4\nu$ Final State in Proton-Proton Collisions with the ATLAS Detector at the LHC*. PhD thesis, Albert-Ludwigs-Universität Freiburg, May, 2016.

[135] G. Hanson et al., *Evidence for Jet Structure in Hadron Production by e^+e^- Annihilation*, Phys. Rev. Lett. **35** (1975) 1609–1612.

[136] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector at the LHC*, Eur. Phys. J. **C76** (2016) 653, `arXiv:1608.03953 [hep-ex]`.

[137] ATLAS Collaboration, *ATLAS TWiki*, `https://twiki.cern.ch/twiki/bin/view/AtlasProtected/LatestRecommendationsElectronIDRun2`. [Online; accessed 17-October-2017].

[138] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data*, Eur. Phys. J. C **74** (2014) 3071, `arXiv:1407.5063 [hep-ex]`.

[139] ATLAS Collaboration, *Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at $\sqrt{s} = 13$ TeV*, ATL-PHYS-PUB-2015-015, 2015, `https://cds.cern.ch/record/2037613`.

[140] ATLAS Collaboration, *ATLAS TWiki*, `https://twiki.cern.ch/twiki/bin/view/AtlasProtected/MissingETSystematicsDerivation`. [Online; accessed 17-October-2017].

[141] J. Bellm et al., *Herwig 7.0/Herwig++ 3.0 release note*, Eur. Phys. J. **C76** (2016) 196, `arXiv:1512.01178 [hep-ph]`.

[142] M. Bahr et al., *Herwig++ Physics and Manual*, Eur. Phys. J. **C58** (2008) 639–707, `arXiv:0803.0883 [hep-ph]`.

[143] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, J. Phys. **G43** (2016) 023001, `arXiv:1510.03865 [hep-ph]`.

[144] ROOT Collaboration Collaboration, K. Cranmer, G. Lewis, L. Moneta, A. Shibata, and W. Verkerke, *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, New York, Jan, 2012.

[145] L. Moneta, K. Cranmer, G. Schott, and W. Verkerke, *The RooStats project*, Proceedings of the 13th International Workshop on Advanced Computing and Analysis Techniques in Physics Research. February 22-27, 2010, Jaipur, India.57, `arXiv:1009.1003 [physics.data-an]`.

[146] ATLAS Collaboration, *ATLAS TWiki*, `https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/WorkspaceMaker`. [Online; accessed 18-October-2017].

[147] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, The European Physical Journal C **71** (2011) 1554.

[148] N. Ruthmann, *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, Albert-Ludwigs-Universität Freiburg, Oct, 2014.

[149] D. L. Rainwater, R. Szalapski, and D. Zeppenfeld, *Probing color singlet exchange in Z + two jet events at the CERN LHC*, Phys. Rev. **D54** (1996) 6680–6689, `arXiv:hep-ph/9605444 [hep-ph]`.

Acknowledgment

First, I would like to thank Prof. Dr. Markus Schumacher for providing me with such an interesting topic for a master thesis and his supervision throughout the year. If any questions regarding particle physics appeared he was willing to help me with his expertise and knowledge.

Further more I would like my immediate supervisor, Dr. Duc Bao Ta for his help with questions regarding physical and technical nature, and helping me survive in the jungle of ATLAS meetings and politics.

Last but not least I would like to express my gratitude to all members of the group to help me feel welcome in the group and providing a pleasant working climate.