ETP-BACHELOR-KA/2019-05

Study of the Impact of Higgs Boson Variables in the Jet-to-Quark Assignment for the Investigation of Higgs Boson Couplings to Top Quarks and Vector Bosons with the CMS Experiment

Untersuchung des Einflusses von Higgs-Boson-Variablen in der Jet-Quark-Zuordnung für die Analyse der Higgs-Boson-Kopplungen an Top-Quarks und Vektorbosonen mit dem CMS-Experiment

Lea Stockmeier

Department of Physics
Karlsruhe Institute of Technology (KIT)

BACHELOR THESIS

Referee: Prof. Dr. Thomas Müller
Co-Referee: Dr. Thorsten Chwalek
Institut für Experimentelle Teilchenphysik

May 20, 2019

Akzeptiert vom ersten Referenten der Bachelorarbeit.
Karlsruhe, den 20. Mai 2019

. . . . . .
(**Prof. Dr. Thomas Müller**)

Ich versichere wahrheitsgemäß, die Arbeit selbstständig angefertigt, alle benutzten Hilfsmittel vollständig und genau angegeben und alles kenntlich gemacht zu haben, was aus Arbeiten anderer unverändert oder mit Abänderungen entnommen wurde.

Karlsruhe, den 20. Mai 2019

. . . . . .

(**Lea Stockmeier**)

Introduction

With the discovery of the Higgs boson in 2012 by the CMS and ATLAS experiments, the last missing particle predicted by the standard model of particle physics (SM) was found. Further Higgs boson studies aim at determining the properties of the discovered particle. Deviations from the SM prediction are a hint for physics beyond the SM.

The analysis presented in this thesis studies different coupling scenarios of the Higgs boson to top quarks and vector bosons. The coupling strength to the Higgs boson is given by g_V for vector bosons and the Yukawa coupling strength y_f for fermions. As the top quark is the heaviest particle of the SM, its Yukawa coupling strength y_t is larger than that of the other fermions. Deviations from the SM predictions are expressed as $\kappa_t = y_t/y_t^{SM}$ and $\kappa_V = g_V/g_V^{SM}$. The different studied coupling scenarios correspond to a specific ratio of κ_t/κ_V.

For the investigation of these coupling scenarios, the production of a Higgs boson in association with a top quark (tHq process), the production of a Higgs boson associated with a W boson and a top quark (tHW process) and the production of a Higgs boson in association with a top quark pair ($t\bar{t}H$ process) are investigated. The tHq and tHW processes, together called tH process, are sensitive to the relative sign of κ_t and κ_V while the $t\bar{t}H$ process provides additional sensitivity to the magnitude of κ_t. The production of a top quark pair ($t\bar{t}$ production) is the most dominant background process. For the discrimination of the signal and background processes, multivariate analysis tools are required. Three different boosted decision trees (BDTs) assign the jets to the final state particles under three different hypotheses, tHq, tHW and $t\bar{t}$ respectively. Subsequently, another BDT classifies the events according to their signal or background process likeness. A final fit on this BDT output determines upper limits on the cross section times branching fraction of the combined tH and $t\bar{t}H$ process for different coupling scenarios.

The event classification BDT utilises global event variables as well as variables depending on the jet-to-quark assignment as input. The jet-to-quark assignment of background events, where no Higgs boson is present, is also performed under the tHq and tHW hypotheses. The use of Higgs-boson-dependent variables in the jet assignment of background events could result in the reconstruction of a Higgs-boson-like signature from arbitrary jets. Hence, background events could be more signal-like decreasing the separation power of the signal classification BDT. Within this thesis, the impact of Higgs-boson-dependent variables in the jet-to-quark assignment on the event classification is studied.

The analysed data set was recorded in 2016 by the Compact Muon Solenoid (CMS) experiment which is located at the Large Hadron Collider (LHC). The centre-of-mass-energy of the proton-proton collisions is $\sqrt{s} = 13\,\text{TeV}$ and the integrated luminosity of the analysed data set is $35.9\,\text{fb}^{-1}$.

The first chapter introduces the relevant theory while chapter 2 explains multivariate

analysis techniques and statistical methods exploited in this thesis. The experimental setup is described in chapter 3. The whole analysis is presented in detail in chapter 4. Chapter 5 summarises the thesis and provides an outlook of further studies related to this thesis.

Contents

1. Theoretical Background

This thesis investigates different coupling scenarios of the Higgs boson. Describing the fundamental particles and their interactions, the standard model of particle physics (SM) is an important theoretical foundation of this analysis. It is briefly described in section 1.1. The relevant processes for the analysis presented in this thesis are explained in section 1.2. Both parts of this chapter are based on reference [1].

1.1. The Standard Model of Particle Physics

The fundamental particles described by the SM (see figure 1.1) are grouped into two categories: Fermions, which have a half-integer spin, and bosons with integer spin. Quarks and leptons are fermions belonging to the SM. They have a spin given by $1/2$, form hadrons, e.g. protons and neutrons, and chemical elements, e.g. the hydrogen atom, and are arranged in three generations. Each generation of quarks is made up of one up-type quark with electric charge $+2/3\,e$ and one down-type quark with electric charge $-1/3\,e$. A charged lepton (electron, muon or tau) with electric charge $-e$ and the corresponding neutral lepton (electron neutrino, muon neutrino or tau neutrino) belong to each generation of leptons. An anti-particle with opposite electric charge exists for all of these fundamental particles.

There are three different types of interaction between these particles: The electromagnetic, weak, and strong interaction. Each interaction has its own gauge bosons (spin 1), which mediate the force between the interacting particles. These gauge bosons couple to the corresponding charge. While the photon couples to the electric charge to mediate the electromagnetic interaction, the W^+, W^-, and Z bosons couple to the weak charge to mediate the weak interaction and the eight gluons couple to the colour charge (blue, red, green and the corresponding anticolours) to mediate the strong interaction. Since gluons and quarks are colour charged, they can interact via the strong force. Especially the gluons interact with each other, so called gluon self-interaction. The leptons, quarks, W bosons and Z bosons carry weak charge, so they can interact through the weak force. The gravitation is not explained by the SM and is negligible for the small scales regarded in particle physics.

Since the gauge invariance forbids mass terms in the Lagrange density (Lagrangian), all masses of the quarks, charged leptons, W bosons and Z bosons are explained by the Higgs mechanism: The electroweak theory combines the electromagnetic and the weak force. Due to spontaneous symmetry breaking, the W and Z bosons are massive and the photon is massless. The spontaneous symmetry breaking is caused by an infinite number of possible ground states for the Higgs potential. One particular of these ground states has to be chosen. The masses of the fermions can be explained by the Lagrangian of the coupling between the Higgs field and the fermion fields. The Yukawa coupling y_f of the

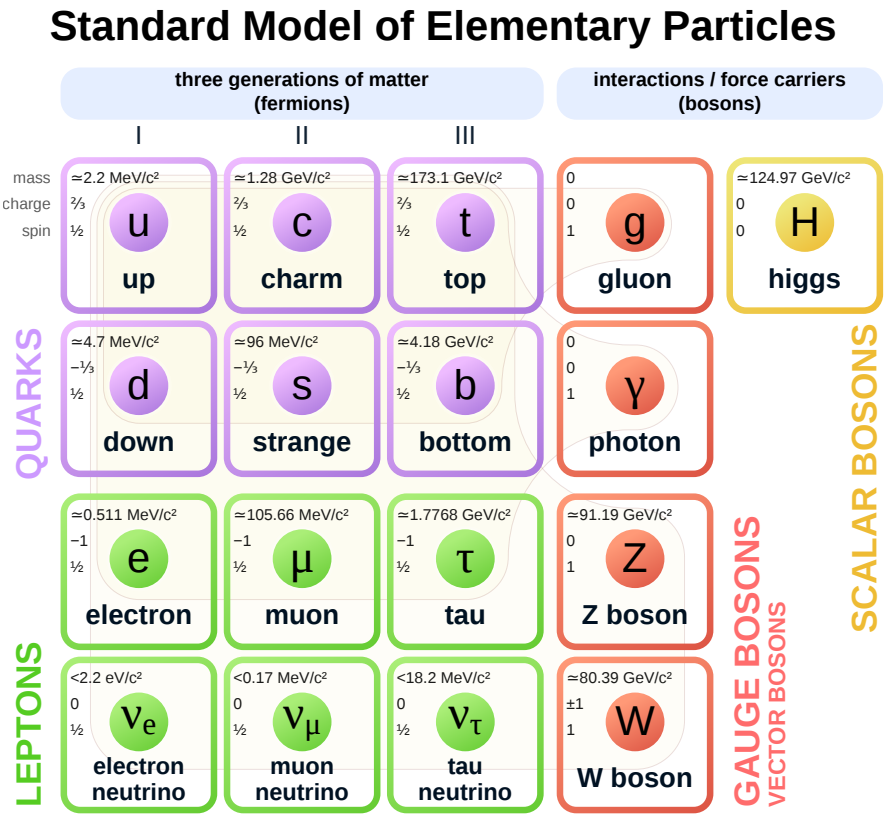


Figure 1.1.: The standard model of particle physics. The particles of the standard model (SM) are grouped into fermions and bosons. Quarks and leptons are fermions. The gauge bosons and the Higgs boson form the bosons of the SM. Figure taken from reference [2].

Higgs boson to fermions depends on the mass of the fermion.

1.2. The Higgs Boson and the Top Quark

In 2012, the Higgs boson was discovered. It is the only scalar boson (spin 0) associated with the SM. Its experimentally determined mass is $m_{\mathrm{H}} = 125.18\,\mathrm{GeV}$ [3]. This section introduces the most relevant production modes of the Higgs boson in proton-proton collisions with huge centre-of-mass-energies.

The top quark forms together with the bottom quark the third quark generation. With a mass of 173.0 GeV [3], it is the heaviest particle belonging to the SM and thus, its Yukawa coupling strength to the Higgs boson is larger than of the other quarks. Therefore, the top quark is important for the production modes of the Higgs boson.

The most important production modes are the gluon-gluon fusion process, the vector boson fusion, the production in association with a vector boson (Higgsstrahlung) and the production in association with top or bottom quark pairs ($\mathrm{t\bar{t}H}$ or $\mathrm{b\bar{b}H}$). The $\mathrm{t\bar{t}H}$ process depends on the magnitude of the Yukawa coupling y_{t} to top quarks. The Feynman diagrams of the production of Higgs bosons in association with single top quarks (tH)

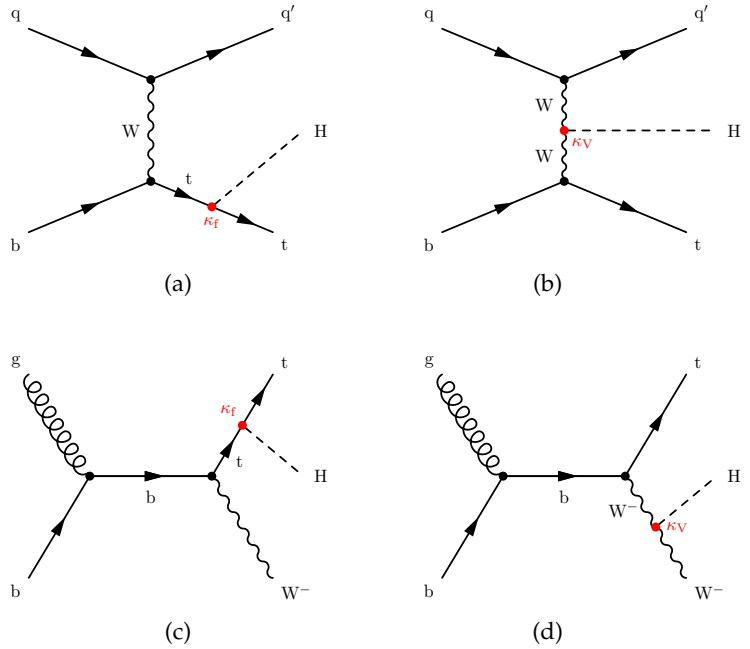


Figure 1.2.: Feynman diagrams of the Higgs boson production in association with a single top quark. The Feynman diagrams of the *t*-channel production of a single top quark in association with a Higgs boson ((a) and (b)) and the production of a Higgs boson together with a W boson ((c) and (d)) are depicted. The Higgs boson can either couple to the top quark ((a) and (c)) or to the W boson ((b) and (d)). Figures taken from reference [4].

are given in figure 1.2. The $t\bar{t}H$ and tH process are researched in this thesis.

The tH process contains of two different production modes: The *t*-channel production of a single top quark in association with a Higgs boson (tHq, see figure 1.2(a) and 1.2(b)) and the production of a Higgs boson together with a W boson and a top quark (tHW, see figure 1.2(c) and 1.2(d)). In both processes, the Higgs boson can either couple to the top quark (see figure 1.2(a) and 1.2(c)) or the W boson (see figure 1.2(b) and 1.2(d)). Thus, the cross sections are given by the interference of both coupling scenarios [5]:

$$\sigma_{tHq}/\text{fb} = 74.25 \cdot (2.63 \cdot \kappa_t^2 + 3.58 \cdot \kappa_V^2 - 5.21 \cdot \kappa_t \kappa_V)\,, \tag{1.1}$$

$$\sigma_{tHW}/\text{fb} = 15.17 \cdot (2.91 \cdot \kappa_t^2 + 2.31 \cdot \kappa_V^2 - 4.22 \cdot \kappa_t \kappa_V)\,. \tag{1.2}$$

They depend on the coupling strength modifiers $\kappa_t = y_t/y_t^{SM}$ and $\kappa_V = g_V/g_V^{SM}$ with the Yukawa coupling strength y_f of fermions. The cross sections of both processes depend on the term $\kappa_t \cdot \kappa_V$ and are therefore sensitive to the relative sign of κ_t and κ_V. The predictions of the SM correspond to $\kappa_t = \kappa_V = 1$ and the inverted top coupling (ITC) case is given by $-\kappa_t = \kappa_V = 1$. The aim of studies like this is to exclude or find different coupling scenarios, which are given by a specific value of κ_t/κ_V. For this, the combined tH and $t\bar{t}H$ process is investigated.

The coupling strength of particles to the Higgs boson scales with the respective particle mass. Since the decay into a top quark pair is forbidden, the dominant decay channel of the Higgs boson is the decay into a pair of bottom quarks. This is the reason why this thesis investigates Higgs bosons decaying into a bottom quark-antiquark pair.

2. Statistical Analysis

For analyses in particle physics like the one presented in this thesis, statistical analysis techniques are essential. This chapter introduces a multivariate analysis technique named boosted decision trees and two statistical methods, referred to as maximum likelihood method and exclusion limit calculation. The whole chapter is based on the references [6] and [7].

2.1. Boosted Decision Trees

Boosted decision trees (BDTs) can be used to classify events into two categories, e.g. signal and background. They are multivariate analysis tools and are used in this thesis for the jet assignment and the classification of events. The implementation of the BDTs is given by the package "Toolkit for MultiVariate Data Analysis" (TMVA) [8] included in the ROOT framework [9].

A simple decision tree (see figure 2.1) is a sequence of binary decisions based on input variables x_i from a vector $\vec{x}$. Binary decision means that there are two possible outcomes of every decision. For every decision, one of the inputvariables x_1, ... x_n is used to decide based on a criterion ($x_i \geq c_j$ or $x_i < c_j$) whether the events are signal- or background-like. The variable x_i and the criterion c_j are chosen in order to maximise the separation gain D between parent node and children nodes. In packages like TMVA, the number of cuts can be chosen. This means that as many grid points in the variable range are used to find the criterion c_j. There are different methods to determine the separation gain. The method taken in this thesis uses the Gini index

$$G = P(1 - P) \tag{2.1}$$

with purity

$$P = \frac{\sum_{i=1}^{n_\mathrm{s}} w_i}{\sum_{i=1}^{n_\mathrm{s}} w_i + \sum_{i=1}^{n_\mathrm{b}} w_i} \tag{2.2}$$

of each branch. The sum includes all signal-like (background-like) weights w_i of events of the branch depending on the upper limit n_s (n_b). Initially, all events are weighted equally with $w_i = 1$. By the use of the Gini index, the separation gain is defined as

$$D = G_{\text{parent node}} - G_{\text{child node 1}} - G_{\text{child node 2}}\,. \tag{2.3}$$

This is repeated until one of the stopping criteria is reached. This can be the maximal permitted depth of the tree, the minimal number of events belonging to a node or the fact that the separation gain is worse than a pre-defined value D_{crit}.

A method to improve the separation between signal and background, to make the method robust against fluctuations and to prevent the learning of statistical fluctuations

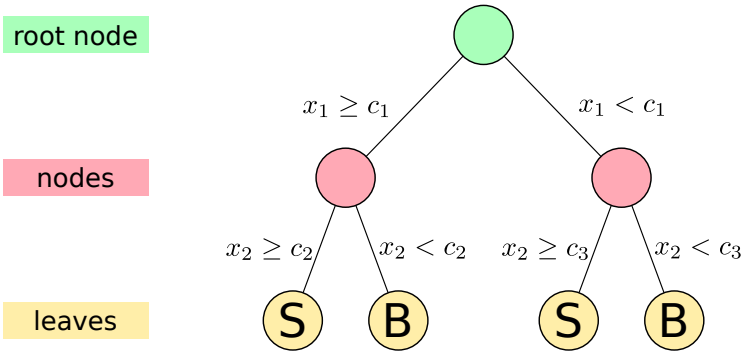


Figure 2.1.: Simple decision tree. Beginning with the root node, the first decision is made with the variable x_1 based on the comparison with the value c_1. The sample is then split into two new nodes. This is repeated with another variable x_2 and other criteria c_2 and c_3. The endnodes are called leaves. In every iteration, the so-called parent node is split along branches into two children nodes. Figure taken from reference [10].

is to use more than one tree. The number of trees M is often chosen between 100 and 1000. While using more than one tree, the events are reweighted after the training of each tree. This procedure is referred to as boosting. The boosting algorithm used in the analysis described in this thesis is the AdaBoost algorithm [11]. The algorithm principle is to reweight the misclassified events. The weights are modified according to

$$w_i \to w_i \cdot e^{\alpha_m S_m(\vec{x})}\,, \tag{2.4}$$

with $S_m(\vec{x}) = 0\ (=1)$ for wrong (correct) assignments by the previous tree. The weight of the previous tree

$$\alpha_m = \ln\left(\frac{1-\epsilon_m}{\epsilon_m}\right) \tag{2.5}$$

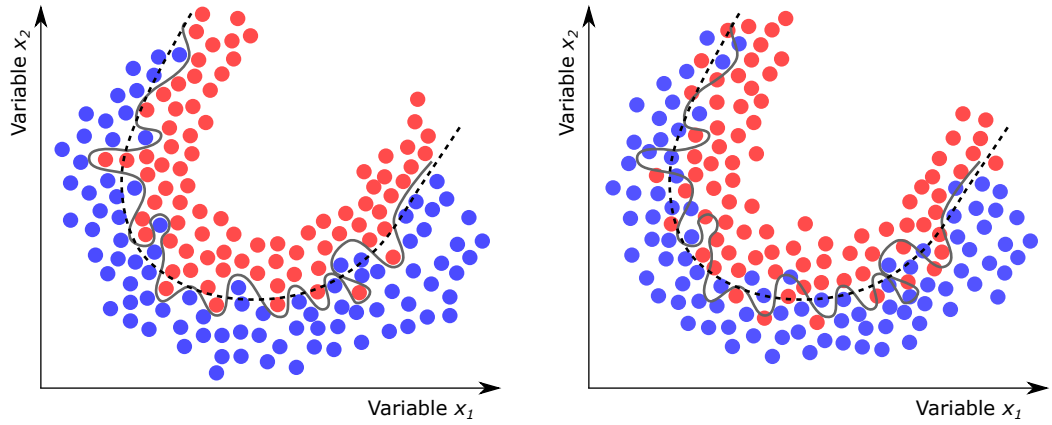
is defined via the misclassification rate ϵ_m of the previous tree. The final BDT output of each event is then estimated by

$$y_{\text{boost}}(\vec{x}) = \frac{1}{M}\sum_{m=1}^{M} \alpha_m^{\beta} \cdot T_m(\vec{x})\,. \tag{2.6}$$

The pre-defined parameter β describes the strength of the boosting and $T_m(\vec{x})$ names the binary output of tree m, i.e. $+1$ for signal and -1 for background.

Generally, multivariate analysis techniques are trained first, then tested and finally applied on samples. It is important that the three samples taken for the three steps are statistically independent. Only if this is given, it is possible to discover if the BDT learned statistical fluctuations of the training sample. This effect is called overtraining or overfitting. Figure 2.2 illustrates what happens if an overtrained BDT is applied to an independent sample. To check whether a BDT is overtrained, the shape of the BDT output of the training sample and a statistically independent test sample is compared. The Kolmogorov-Smirnov (KS) test is the method for comparing two shapes used in this analysis.

After the training of a BDT, it has to be checked how good it can differentiate between signal and background. For this aim, the receiver operating characteristic curve (ROC



(a) (b)

Figure 2.2.: Illustration of overtraining. The separation power of the BDT is better on the training sample (a) than on the test sample (b) because the BDT learned the statistical fluctuations of the training sample (solid lines). The training performing equally on both samples corresponds to a not overtrained BDT (dashed lines). Figure adapted and modified from reference [12].

curve) is useful. It can be calculated by scanning the whole range of the BDT output. For every value of the BDT output, the true-positive-rate is plotted against the false-positive-rate. The goal is, to maximise the true-positive-rate and to minimise the false-positive-rate. This is given for one value of the BDT output, so-called maximum working point, which is then used to separate between signal and background. The area under the ROC curve is defined as the integral of the ROC curve. Its maximal possible value is one and the bigger its value is, the better is the separation power of the trained BDT.

The input variables can be ranked by their importance in the training. This ranking is useful to check if the separation power is the same for less variables without the least important. The goal of training a BDT is to maximise the separation power and to minimise the overtraining and the number of input variables. This can be reached by adapting the parameters explained previously in this section.

2.2. Maximum Likelihood Method

The maximum likelihood method is a statistical method which can be utilised to find the best configuration of the parameters of a function describing the measurement. This method is based on maximising the likelihood function

$$L(\vec{a}) = \prod_{i=1}^{n} f(\vec{x}_i|\vec{a}) \leq L(\hat{\vec{a}})\,. \tag{2.7}$$

$\vec{x}_i$ denotes the n measured variables and $\vec{a}$ the m parameters to be determined. $f(\vec{x}_i|\vec{a})$ is the probability density function depending on the variable $\vec{x}_i$. The configuration of parameters $\vec{a}$ describing the measurement best is then $\hat{\vec{a}}$.

For binned distributions the likelihood function is given by

$$L(\vec{a}) = \prod_{i=1}^{m} P(n_i|\lambda_i) = \prod_{i=1}^{m} \frac{\lambda_i^{n_i}}{n_i!} \cdot \mathrm{e}^{-\lambda_i}\,, \tag{2.8}$$

where $i \in \{1, ...\, m\}$ is the bin number, n_i the number of observed entries in bin i and $P(n_i|\lambda_i)$ the poisson probability for bin i. The expected number of events λ_i is defined as

$$\lambda_i = \lambda_i(\mu, \vec{\theta}) = \mu \cdot \lambda_{\mathrm{s},i}(\vec{\theta}) + \lambda_{\mathrm{b},i}(\vec{\theta})\,, \tag{2.9}$$

where $\lambda_{\mathrm{s},i}$ is the expected number of signal events and $\lambda_{\mathrm{b},i}$ the expected number of background events in bin i. The expected number of signal events is multiplied by the signal strength $\mu = {}^{\sigma}/_{\sigma_{\mathrm{SM}}}$, which depends on the physical model. The nuisance parameters $\vec{\theta}$ include the shape and rate uncertainties in the likelihood function.

It is more comfortable to use the negative log-likelihood-function

$$-\ln[L(\vec{a})] = -\ln\left[\prod_{i=1}^{n} f(\vec{x}_i|\vec{a})\right] = -\sum_{i=1}^{n} \ln\left[f(\vec{x}_i|\vec{a})\right] \tag{2.10}$$

since there is a sum instead of a product and a minimum instead of a maximum can be evaluated.

2.3. Exclusion Limit Calculation

Due to the fact that the analysed processes have small predicted cross-sections and similar final states as other processes, they are overlain by many different background processes. Therefore, it is not possible to measure the cross-section of the signal process directly without using the exclusion limit calculation.

The exclusion limit calculation used in this analysis is the $\mathrm{CL_s}$ method [13]. The aim is to exclude one of two hypotheses, i.e. the signal+background hypothesis (s+b) and the background only hypothesis (b). The signal strength modifier of the background hypothesis is $\mu = 0$ as there is no signal expected ($\sigma = 0$). This is the reason why the background only hypothesis is also named null hypothesis.

The signal+background hypothesis can be excluded if $\mathrm{CL_s} \leq \alpha$ with the so-called pre-defined significance level α. A value of $\alpha = 0.05$ is equivalent to a 95 % confidence level which means that the analysed signal does not exist as predicted from theory with a probability of 95 %. As normal in particle physics, in this analysis $\alpha = 0.05$ is taken.

In this analysis, upper limits on μ are calculated by estimating the confidence level of signal $\mathrm{CL_s}$ for different values of μ. The smallest value of μ that can reject the null hypothesis is the upper limit of μ.

A huge discrepancy between expected and observed limit can be a hint for new physics that was not considered yet or for no adequate treatment of the expected uncertainties.

This thesis uses the COMBINE package [14] for the maximum likelihood method and the exclusion limit calculation based on the RooStats toolkit [15].

3. Experimental Setup

The aim of large experiments in particle physics with high energy of particles is to reveal the physics beyond the standard model and to measure properties of standard model processes. This chapter describes in section 3.1 the Large Hadron Collider (LHC) and in section 3.2 the Compact Muon Solenoid (CMS) experiment. Section 3.3 provides an overview of the simulation and reconstruction of events.

3.1. The Large Hadron Collider

The Large Hadron Collider [16, 17] at the European Organisation for Nuclear Research (CERN, french: Conseil Européen pour la Recherche Nucléaire) is located about 100 m under ground near Geneva in Switzerland and France. With a circumference of 27 km, it is the largest particle collider of the world. Together with smaller pre-accelerators, the LHC was constructed to collide protons with a centre-of-mass-energy of $\sqrt{s} = 14\,\text{TeV}$. Another type of application is to accelerate and collide heavy ions (lead (Pb) ions) with an energy of 2.8 TeV per nucleon.

There are four points where the particle beams collide. To detect these collisions, four particle detectors, called experiments, were built. The experiments have different goals, which are explained briefly in the following:

A Large Ion Collider Experiment (ALICE) ALICE investigates the properties of quark-gluon plasma in heavy ion collisions. Further information can be found in reference [18].

Large Hadron Collider beauty (LHCb) The LHCb measures CP violation and rare decays of B hadrons. For further information have a look at reference [19].

A Toroidal LHC Apparatus (ATLAS) ATLAS tests the QCD, electroweak interactions and flavour physics. The search for the Higgs boson and the investigation of the properties of the Higgs boson are other tasks. For further information see reference [20].

Compact Muon Solenoid (CMS) The CMS experiment has the same physics goals as ATLAS. Due to their different experimental setup, both experiments yield independent measurements. In 2012, CMS and ATLAS discovered the Higgs boson [21, 22].

In this thesis, the data set recorded in 2016 with the CMS experiment is analysed. The centre-of-mass-energy of the colliding particles is $\sqrt{s} = 13\,\text{TeV}$ and the integrated luminosity of the data set is $37.8\,\text{fb}^{-1}$ [23] of which $35.9\,\text{fb}^{-1}$ are usable for this analysis.

3.2. The Compact Muon Solenoid Experiment

The CMS experiment [24] is located in the north of the LHC at Point 5. The CMS detector has a cylindrical shape with a length of 21 m, a diameter of 15 m and a weight of 14,000 t [25]. The point of the beam pipe, in which the collisions happen, is located in the centre of the detector. This collision point is the origin of the coordinate system used to describe the traces of the particles. The x-axis points toward the centre of the LHC, the y-axis points vertically upward and the z-axis points in direction of the proton beam circling counterclockwise through the LHC when viewed from above ground. Cylindrical coordinates with azimuthal angle ϕ and polar angle θ are used. ϕ is located in the x-y-plane and measured from the x-axis. The radial coordinate $r = \sqrt{x^2 + y^2}$ and z-axis define the plane in which θ is measured beginning on the z-axis. Instead of θ, the pseudorapidity

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right] \tag{3.1}$$

is often used. To describe the positions of two particles in relation to each other, the spatial distance

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \tag{3.2}$$

is commonly used [26]. In contrast to the momentum p_z in beam direction, the transverse momentum

$$p_\mathrm{T} = \sqrt{p_x^2 + p_y^2} \tag{3.3}$$

of the accelerated particles is known and zero before the collision. This is the reason why it is often used as a variable describing the particles in the analyses based on data from the CMS experiment.

Figure 3.1 shows a sector of the CMS detector with its subdetectors. In the following they are described from inside out.

Tracking System The aim of the tracking system is to measure the trajectories of charged particles. It is made up of two parts, the pixel detector and the silicon strip detector. The principle of this detector is that charged particles generate an electron-hole-pair in semiconductors like silicon. In the electric field of the depletion zone, they separate and can be measured as a current. Due to the detector being split in pixels respectively strips, the necessary spatial resolution is achieved.

Electromagnetic Calorimeter (ECAL) The ECAL measures the energy of electrons, positrons and photons. It is composed of lead tungstate ($PbWO_4$) crystals. This material absorbs the measured particles completely by creating electromagnetic showers due to bremsstrahlung and pair production. The intensity of the resulting scintillation light is measured with photo-diodes and is proportional to the energy of the primary particles.

Hadron Calorimeter (HCAL) The HCAL is mounted for the energy measurement of hadrons which create a shower in this part of the CMS detector. Together with the ECAL, the HCAL measures nearly the whole energy of an event. The energy of neutrinos is not measured and the energy of muons is measured partially because

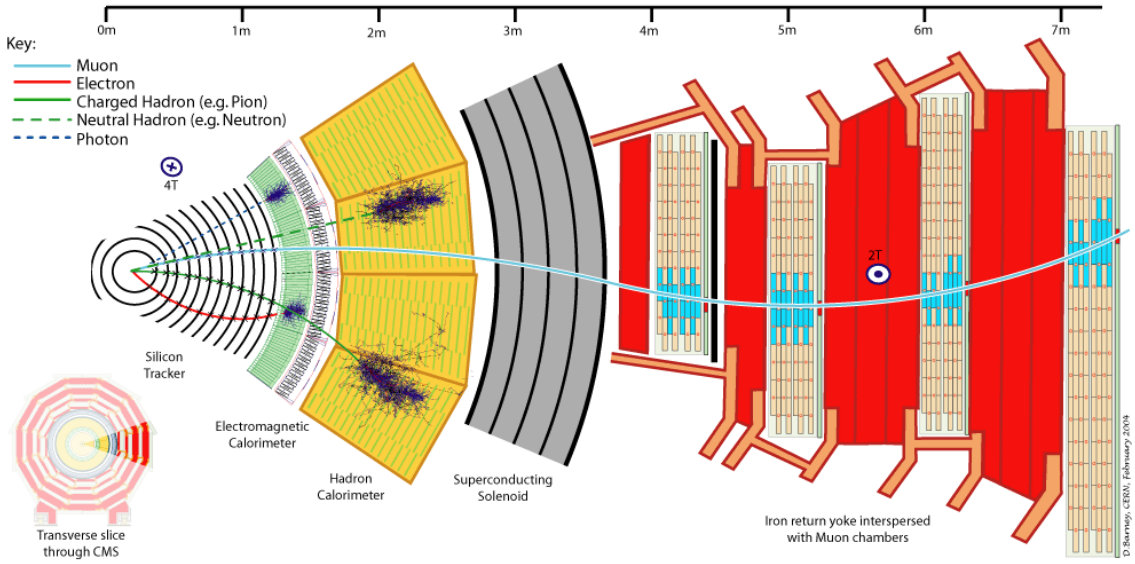


Figure 3.1.: The CMS detector. The figure shows a sector of the CMS detector with distance of the layers to the beam pipe. The different subdetectors beginning close to the beam pipe are the tracking system, the ECAL, the HCAL, the superconducting solenoid and the iron return yoke interspersed with muon chambers. Tracks of different particles and their interaction with the detector are depicted as well. The inset shows a cross section of the whole CMS detector with an indication of the shown sector. Figure taken from reference [27].

their interaction with this part of the detector is very low. If the final state includes only one neutrino, this fact can be used to determine the energy and transverse momentum of this neutrino as missing transverse energy of this event.

Superconducting Solenoid The next layer is a solenoid with a central magnetic flux density of 3.8 T [28]. The direction of its magnetic field is opposite inside and outside the solenoid. A steel return yoke outside ensures homogeneity of the magnetic field. The magnetic field implicates that charged particles have a curved trace in the detector. This, together with the tracking system, is used to measure the transverse momentum of these particles.

Muon System Together with the superconducting solenoid, this part of the detector is eponymous for the whole experiment. Measurements of processes with neutrinos in the final state benefit from the muon system. In contrast to other charged particles, the muons reach the outside part of the detector because their loss of energy by interacting with the detector is smaller than for the other particles. The opposite directions of the magnetic field cause s-curved tracks of muons (see figure 3.1). The muon system contains a drift tube system, cathode strip chambers and a resistive plate chamber system.

3.3. Simulation and Reconstruction of Events

The processes investigated in this theses occur in proton-proton collisions. Due to the fact that protons are not elementary particles, the constituents of the proton (three valence quarks (u, u, d), sea quarks and gluons), so-called partons, interact with each other. The difficulty is that the momentum of a proton is distributed to its partons and the values of these parton momenta are not known. Nevertheless, the parton distribution functions (PDF) can be measured. They describe the probability of the occurrence of a parton with a specific momentum in the proton.

A proton-proton scattering process consists of different parts. Since the energy of the colliding protons is large, the collision can be described as an interaction of free partons, the so-called hard scattering process. It is described by Feynman diagrams and its probability is estimated with perturbation theory. Other interactions with less energy are collectively referred to as underlying event (UE). During the interaction process gluons and photons are radiated. This phenomenon is called initial (final) state radiation depending on whether the radiation occurs in the initial (final) state. These emitted particles can split into even more particles, which form the so-called parton showers. Since gluons carry colour charge and because of the confinement, they have to form colour-neutral states (hadronisation).

The event simulation is done with Monte Carlo (MC) event generators. They are able to start with the initial protons and to simulate how the processes and final states would look like in case of a certain theory. In this thesis, the combination of different simulation software (MadGraph5 [29], aMC@NLO [30], Powheg [31] and Pythia [32]) is used due to the fact that every generator is specialised in simulation of a specific part of proton-proton collisions. In reality, the occurring particles cannot be measured without their interaction with the detector. Due to this fact it is essential to simulate the interaction of the simulated particles with every part of the detector. For the detector simulation the software Geant4 [33] is used.

After the event simulation, the events can be reconstructed. To find the tracks of charged particles in the tracking system, the Combinatorial Track Finder (CTF) [34] is used. The Particle-Flow (PF) algorithm [35] then links the information of the different subdetectors. This makes the identification of physical objects possible. The narrow bundles of particles resulting from hadronisation of one primary quark or gluon are called jets. For this thesis, it is important to identify jets resulting from bottom quarks (b jets) because processes with an Higgs bosons decaying into two bottom quarks are investigated. For the identification of b jets the Combined Secondary Vertex (CSV) algorithm [36] is used. Neutrinos are the only particles which do not interact with the detector. They can be measured indirectly with the use of the missing transverse momentum

$$p_{\mathrm{T}}^{\mathrm{miss}} = \left| -\sum_{i=1}^{n} \vec{p}_{\mathrm{T},i} \right| . \tag{3.4}$$

$\vec{p}_{\mathrm{T},i}$ is the transverse momentum of the reconstructed particle i and n is the total number of reconstructed particles.

4. Investigation of Higgs Boson Couplings to Top Quarks and W Bosons

The study of Higgs boson couplings to top quarks and W bosons investigates processes depending on the Yukawa coupling y_t and the coupling g_V of the Higgs boson to vector bosons. This study considers the production of a Higgs boson in association with a top quark (tH) and the production of a Higgs boson in association with a top quark-antiquark pair ($t\bar{t}H$) as signal processes. The tH process is sensitive to the relative sign of κ_t and κ_V and the $t\bar{t}H$ process provides sensitivity to the magnitude of κ_t.

From the different possible decay modes of the Higgs boson, the decay into a bottom quark-antiquark pair is targeted in this analysis. Another requirement for the decay of the final state particles is that exactly one W boson decays leptonically.

The analysed data set, corresponding to an integrated luminosity of $35.9\,\text{fb}^{-1}$, was recorded in 2016 with the CMS experiment during proton-proton collisions at a centre-of-mass-energy of $\sqrt{s} = 13\,\text{TeV}$.

This thesis analyses the impact of Higgs-boson-dependent variables in the jet assignment on the analysis. It determines the improvement of the analysis sensitivity accomplished with the use of Higgs-boson-dependent variables in the jet assignment and researches if the jet assignment without using Higgs-boson-dependent variables is able to improve the classification that gets jet assignment dependent variables as input. The general aim of studies like this is to improve the sensitivity of the analysis which means that the upper limits on the cross section times branching fraction get smaller.

In section 4.1 and 4.2, the signal and background processes of the analysis are presented, while section 4.3 provides an overview of the analysis strategy. The jet assignment is explained in section 4.4. Section 4.5 describes the classification of the events. The uncertainties taken into account in this thesis are briefly introduced in section 4.6 and the results are presented in section 4.7.

4.1. Signal Processes

As this analysis searches for the production of a Higgs boson in association with a single top quark (tH), there are two signal processes considered: the t-channel single top quark production associated with a Higgs boson (tHq) and the production of a Higgs boson together with a W boson and a top quark (tHW).

The production of a Higgs boson together with a pair of top quarks ($t\bar{t}H$) is sensitive to the magnitude of κ_t. Due to this fact it is considered as signal in the investigation of different coupling scenarios. Nevertheless, it is treated as background in the search for tH production.

Each of the three signal processes is explained in the following and representative

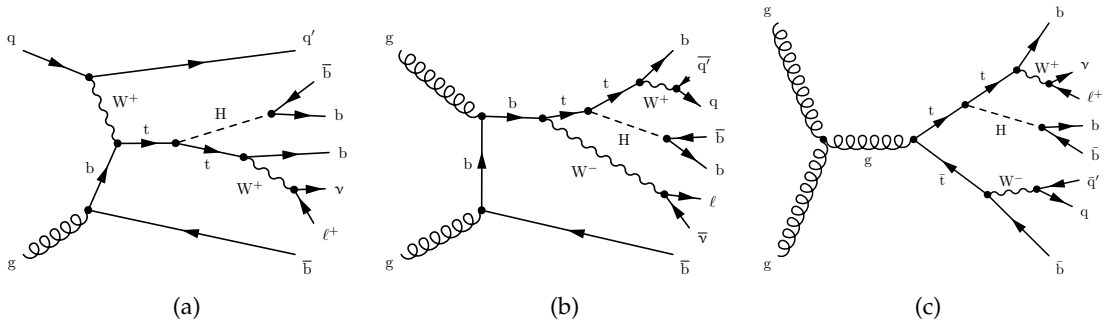


Figure 4.1.: Extended Feynman diagrams of the signal processes. The figure shows one possible Feynman diagram for tHq (a), tHW (b) and $t\bar{t}H$ (c), which are the signal processes. They are extended by the decays of the produced particles.

Feynman diagrams, including the decays of the final state particles, are given in figure 4.1.

Production of a Higgs Boson Associated with a Top Quark The initial state is given by a light-flavoured quark and a gluon (see figure 4.1(a)). The gluon decays into a pair of bottom quarks. One of them interacts with the light-flavoured quark by exchanging a W boson. The produced top quark can then couple to a Higgs boson. Another possibility is that the Higgs boson couples to the W boson. The Higgs boson is assumed to decay into a pair of bottom quarks and the top quark decays into a bottom quark and a W^+ boson. Due to the fact that this W^+ boson is assumed to decay leptonically, the final state is given by: one light-flavoured quark, four bottom quarks, one charged lepton and the corresponding neutrino. Since the bottom quark, which resulted from the gluon splitting, often has a low p_T and is emitted in forward direction, its jet can not always be b-tagged. Hence, three or four b-tagged jets are allowed for this process in the event selection.

Production of a Higgs Boson Associated with a W Boson and a Top Quark The initial state of this process includes two gluons that become a pair of bottom quarks (see figure 4.1(b)). One of the bottom quarks radiates a W^- boson and changes its flavour to top. The occurring Higgs boson couples either to the top quark or to the W boson. As considered in the analysis, one of the W bosons is assumed to decay leptonically and the other one hadronically. The final state is given by two light-flavoured quarks, four bottom quarks, one charged lepton and the corresponding neutrino.

Production of a Higgs Boson in Association with a Top Quark Pair This process corresponds to the top quark pair production with a Higgs boson radiated from one of the top quarks. Depending on the exact process (gluon-gluon fusion (see figure 4.1(c)) or quark-antiquark annihilation) different initial states are possible. As before, the Higgs boson is assumed to decay into two bottom quarks, one of the two W bosons is assumed to decay leptonically and the other one is assumed to decay hadronically.

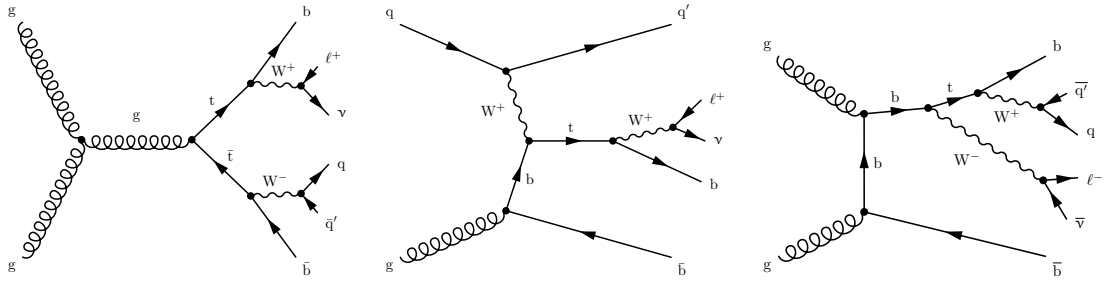


Figure 4.2.: Extended Feynman diagrams of dominant background processes. The figure shows one possible Feynman diagram for $t\bar{t}$ production (a), t-channel production of single top quarks (b) and tW production (c), which are dominant background processes. They are extended by the decays of the produced particles.

4.2. Background Processes

In this section, the dominant background processes are explained and the minor and negligible background processes are listed.

Top Quark Pair Production The gluon-gluon fusion (see figure 4.2(a)) and the quark-antiquark annihilation are two modes of the top quark pair production ($t\bar{t}$ process). Additional jets occur from initial or final state radiation. The possible additional particles are: light-flavoured quarks or gluons ($t\bar{t} + \text{LF}$), two charm quarks ($t\bar{t} + c\bar{c}$), two bottom quarks ($t\bar{t} + b\bar{b}$), two bottom quarks and only one reconstructed as b jet ($t\bar{t} + b$) or two bottom quarks reconstructed in the same jet ($t\bar{t} + 2b$). These five resulting categories are treated separately to reduce the systematic uncertainties.

Single Top Quark Production From the electroweak production of single top quarks (t), the t-channel production (see figure 4.2(b)) and the production in association with a W boson (tW, see figure 4.2(c)) are considered. If there is an additional radiation of a bottom quark pair, this process has the same final state as the signal processes. Therefore, the separation of both processes from the signal process is very difficult.

Minor and Negligible Backgrounds The minor backgrounds (top quark pair production in association with a vector boson, single top quark production in association with a Z boson and Z boson production in association with jets, the so-called Drell-Yan process) are treated separately in the analysis but are shown collectively as "Misc" in the resulting plots.

The diboson production, the W boson production in association with jets and the QCD multijet events are negligible background processes.

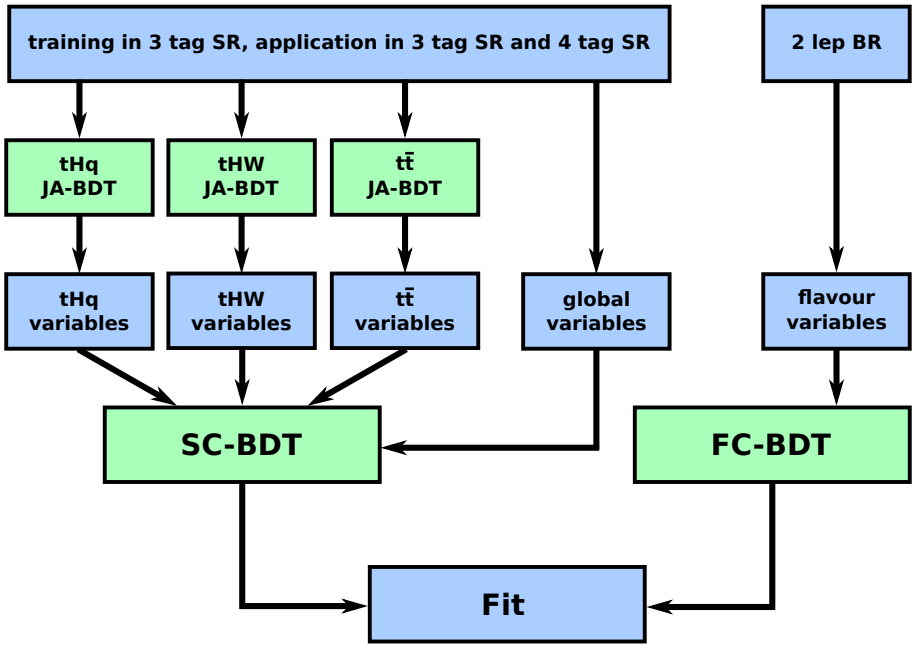


Figure 4.3.: Analysis workflow. Jet assignment BDTs (JA-BDTs) are trained under the tHq, tHW and $t\bar{t}$ hypotheses in the 3 tag SR and applied on the 3 tag SR and 4 tag SR. A signal classification BDT (SC-BDT) uses global event variables and JA-BDT dependent variables for the training. In the dileptonic background region (2 lep BR), a dedicated flavour classification BDT (FC-BDT) is trained and applied. The fit to the SC-BDT output and FC-BDT output derives the resulting limits. Figure adapted and modified from reference [10].

4.3. Analysis Strategy

This section introduces the search strategy of the analysis presented in this thesis. The workflow explained in the following is illustrated in figure 4.3.

As explained in the previous chapters, the final states of the signal processes have characteristic features that are helpful to select them. The challenge is that the final states of the background processes are very similar to those of the signal processes. The cross section of the main background processes is also larger than of the signal process.

The final states of the signal processes are given by one charged lepton, the corresponding neutrino, and several quarks. Some of the quarks are bottom quarks. The jets of bottom quarks can be identified via multivariate analysis methods (b tagging). As a result, the events are sorted by the number of b-tagged jets into categories. Only events with exactly one lepton, $p_{\mathrm{T}}^{\mathrm{miss}} > 45\,\mathrm{GeV}$ ($p_{\mathrm{T}}^{\mathrm{miss}} > 35\,\mathrm{GeV}$) for events with an electron (muon) and jets are selected. The 3 tag signal region (SR) contains events with three b-tagged jets and the 4 tag SR contains events with four b-tagged jets. The events of both regions contain additionally at least one untagged jet. Together they are called single lepton regions.

For the discrimination between signal and background events, jet assignment BDTs (JA-BDTs) are used. They assign the measured jets to the final state quarks of the events under three different event hypotheses: tHq, tHW, and $t\bar{t}$. Using several variables depending on those jet assignments and global event variables, a signal classification BDT

(SC-BDT) combines the variables into a single BDT output. The JA-BDTs and the SC-BDT are trained on simulated samples corresponding to the 3 tag SR and applied on samples belonging to the 3 tag SR and the 4 tag SR.

The $t\bar{t}$ process is the most important background. To understand its impact on the analysis better and to reduce the systematic uncertainties corresponding to the $t\bar{t}$ process, there is another region containing events with two leptons, the so-called dileptonic background region (2 lep BR). In this region, a flavour classification BDT (FC-BDT) is trained and applied.

The last step of the analysis is the combined fit on the SC-BDT output and the FC-BDT output. This fit determines the upper limits on the cross sections times branching fraction for the tH and $t\bar{t}H$ processes for different coupling scenarios specified by the ratio κ_t/κ_V.

4.4. Jet Assignment

The aim of the jet assignment is to determine variables of the event such as the mass of the Higgs boson. Those variables depend on the process of the event and on the assignment of the jets to the final state particles. This assignment is done based on the jet assignment BDTs (JA-BDTs), which assign the measured jets to the final state particles under a specific hypothesis (tHq, tHW, or $t\bar{t}$).

The JA-BDTs are trained in the 3 tag SR. To make sure that they are trained and applied on statistically independent samples, the 3 tag SR is split into two parts including an equal number of events. One of the samples contains the events with an even event number and the other contains the events with an odd event number. These samples are called sample A and sample B in the following. For each hypothesis two JA-BDTs with the same BDT configurations and input variables are trained. The one that is trained on sample A is applied on sample B and the other way around. For the training, sample A and B are split into a training and a test sample including 80 % and 20 % of the respective sample. This procedure allows to use the whole simulated and recorded samples for the analysis. In figure 4.4, the method of training and application of the JA-BDTs is illustrated. The JA-BDTs are applied on the 3 tag SR and 4 tag SR of the MC samples, the data set and the samples representing the systematic uncertainties. If not stated otherwise, the plots contain the events belonging to sample A and B.

In the training, the correct assignment is treated as the signal process which results in a larger BDT output and the wrong assignments are treated as background processes resulting in a lower BDT output. During the application of the JA-BDTs, all possible assignments are calculated and the one with the highest JA-BDT output is then used to calculate the variables which depend on the jet assignment and the respective hypothesis.

The training is performed for the ITC case ($\kappa_t = -1$, $\kappa_V = +1$). The other coupling scenarios can be calculated using a reweighting procedure.

In contrast to the reference analysis, the analysis presented in this thesis uses no Higgs-boson-dependent variables as input of the JA-BDTs. In the subsections 4.4.1, 4.4.2 and 4.4.3 the details and these innovations of the jet assignment under the tHq, tHW, and $t\bar{t}$ hypothesis are explained. An improvement to reduce fluctuations in the background hypothesis is explained in subsection 4.4.4.

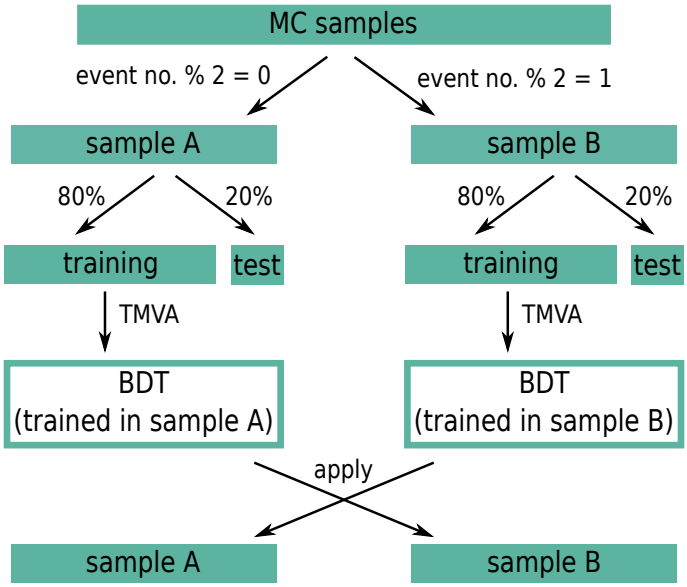


Figure 4.4.: Illustration of the JA-BDT training, test and application. The MC samples are split into sample A and sample B depending on the event number. In each set of events, an independent BDT is trained and tested and afterwards applied on the other set of events.

4.4.1. Jet Assignment under the tHq Hypothesis

The jet assignment under the tHq hypothesis assigns the three b-tagged jets and the one jet according to the light-flavoured quark to the final state quarks. Since more than four jets per event can be reconstructed, the number of possible assignments per event

$$N_A = \frac{N_j!}{(N_j - 4)!} \tag{4.1}$$

depends on the number of assignable jets N_j. To get less possibilities, there are more requirements for the assignments: The jets assigned to the bottom quark need to be central corresponding to $|\eta| < 2.4$, the jet assigned to the light-flavoured quark must not be b-tagged and $\Delta R(q,j) < 0.3$ has be satisfied. If no assignment fulfils all criteria, there is no correct assignment given for the respective event. If more than one assignment is possible, the one with the smallest sum of spatial distances

$$\Delta\Pi = \sum_{(q,j)} \Delta R(q,j) \tag{4.2}$$

between the quark-to-jet assignments (q,j) in the η-ϕ-plane is chosen as the correct assignment. All other assignments are wrong and taken into account as background in the training. Weights depending on the number of wrong assignments of the respective event make sure that all events are weighted equally.

The configuration of the tHq JA-BDT used in this thesis can be found in table 4.1. The input variables are listed in table 4.2. In contrast to further analyses, no variables depending on the reconstructed Higgs boson, e.g. the mass or the transverse momentum of the reconstructed Higgs boson, are used. The distributions of the tHq JA-BDT output for correct and wrong assignment are shown in figure 4.5. Since the distributions of the

Table 4.1.: Configuration of the JA-BDTs. The table contains the configuration of the JA-BDTs used for the jet assignment under the tHq, tHW and $t\bar{t}$ hypothesis. The percentages are based on the total number of events.

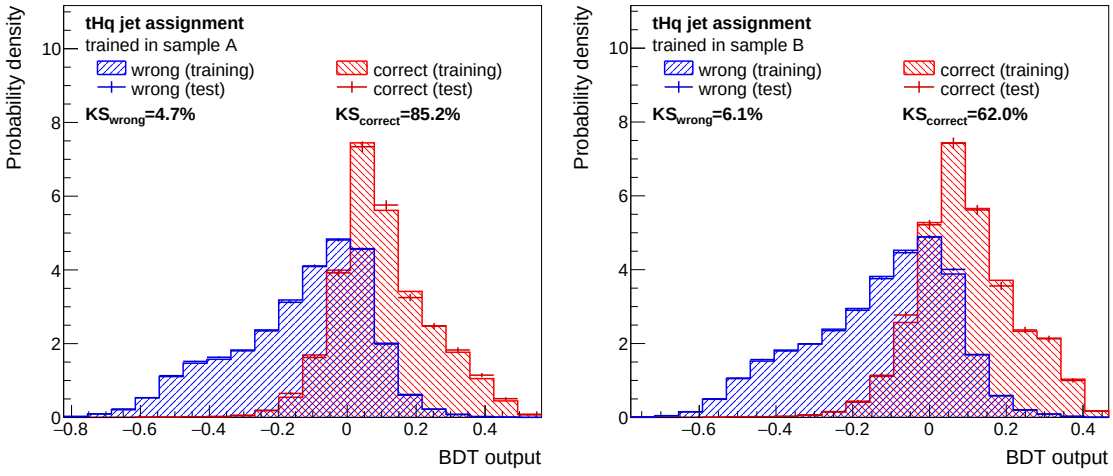
Parameter	tHq JA-BDT	tHW JA-BDT	$t\bar{t}$ JA-BDT
`NTrees`	700	600	400
`MinNodeSize`	1 %	1 %	1 %
`MaxDepth`	3	4	3
`BoostType`	AdaBoost	AdaBoost	AdaBoost
`AdaBoostBeta`	0.3	0.3	0.3
`nCuts`	10	10	20
`SeparationType`	GiniIndex	GiniIndex	GiniIndex

training and test samples are similar, no overtraining is observed. In figure 4.6, the shape of the best tHq JA-BDT response for the ITC scenario in the 3 tag SR is given. Agreement of simulation and data is determined.

The tHq JA-BDT configuration and the input variables taken in the reference analysis are given in table A.1 and A.2 in the appendix A.

Table 4.2.: Input variables of the tHq JA-BDT. The variables are ranked by their importance in the training performed in sample A. Variations of the ranking in sample A and B begin with the variable ranked 6$^{\text{th}}$ in the training.

Variable	Description
$\Delta R(\mathrm{b_t}, l)$	ΔR between the jet assigned to the bottom quark from the top quark decay and the lepton
$\log m(\mathrm{t})$	Invariant mass of the reconstructed top quark
$p_{\mathrm{T}}(\mathrm{t})$	Transverse momentum of the reconstructed top quark
$\cos\theta(\mathrm{b_t}, l)$	Cosine of the angle between the jet assigned to the bottom quark from the top quark decay and the lepton
$\lvert\eta(\mathrm{b_t})\rvert$	Absolute pseudorapidity of the jet assigned to the bottom quark from the top quark decay
$p_{\mathrm{T}}(\mathrm{light\,jet})$	Transverse momentum of the light forward jet
$\lvert\eta(\mathrm{light\,jet}) - \eta(\mathrm{b_t})\rvert$	Absolute difference of the pseudorapidity of the light forward jet and the bottom quark from the top quark decay
$\lvert\eta(\mathrm{light\,jet})\rvert$	Absolute pseudorapidity of the light forward jet
$\Delta E(\mathrm{light\,jet}, \mathrm{b_t})$	Jet energy difference of the light forward jet and the jet assigned to the bottom quark from the top quark decay
$\lvert\eta(\mathrm{t})\rvert$	Absolute pseudorapidity of the reconstructed top quark
$\mathrm{CSV(b_t)}$	Output of the b tagging discriminant for the jet assigned to the bottom quark from the top quark decay



(a) (b)

Figure 4.5.: Distribution of the tHq JA-BDT output. The figure shows the distribution of the tHq JA-BDT output trained in sample A (a) and trained in sample B (b) for wrong (blue) and correct (red) jet assignments. The distributions of the training samples (dashed) and test samples (markers) look similar. No overtraining is observed.

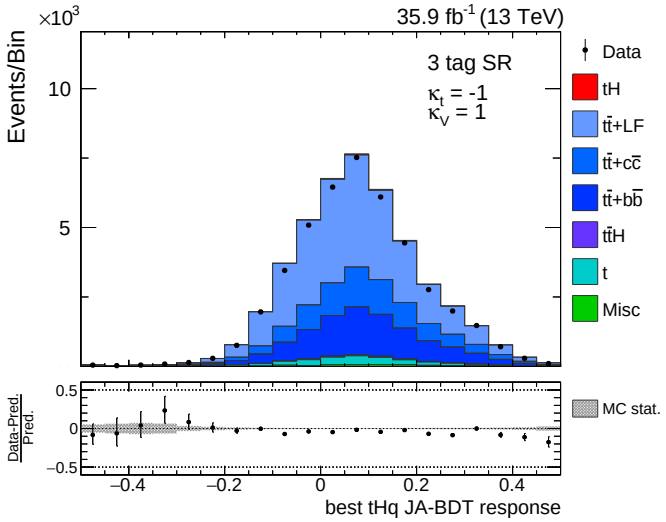


Figure 4.6.: Distribution of the best tHq JA-BDT response. The distribution of the best tHq JA-BDT response for the ITC scenario is shown for the 3 tag SR. Agreement of simulation and data is observed.

4.4.2. Jet Assignment under the tHW Hypothesis

The jet assignment under the tHW hypothesis matches the three bottom quarks and the two light-flavoured quarks of the final state to the reconstructed jets. The number of possible assignments is given by

$$N_{\mathrm{A}} = 2 \cdot \frac{N_{\mathrm{j}}!}{(N_{\mathrm{j}} - 5)!} \,. \tag{4.3}$$

The factor two results from the fact that one of the two W bosons is assumed to decay leptonically. Since the single lepton region contains events with just one or more not b-tagged jets, there can be less jets than necessary. In the training, just events with two or more not b-tagged jets are used. The conditions for the correct assignment are the same as in the jet assignment under the tHq hypothesis. If more than one assignment satisfies the conditions of correct assignment, the one with the smallest sum

$$\Delta\Pi = \sum_{(\mathrm{q},\mathrm{j})} (\Delta R(\mathrm{q},\mathrm{j})) + \Delta R(\mathrm{t}_{\mathrm{gen}}, \mathrm{t}_{\mathrm{rec}}) \tag{4.4}$$

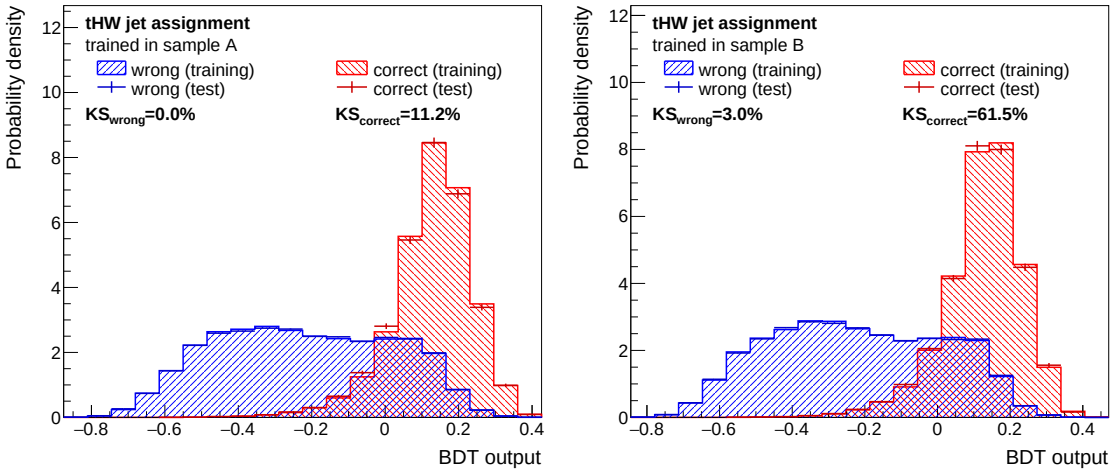
is taken as the correct assignment. $\Delta R(\mathrm{t}_{\mathrm{gen}}, \mathrm{t}_{\mathrm{rec}})$ is the distance of the simulated and the reconstructed top quark in the η-ϕ-plane.

In the application of the tHW JA-BDT, the variables of events with no possible assignment are set to a global default value of $-99,999$ and the tHW JA-BDT output of these events is set to -1.

The training parameters of the tHW JA-BDT are listed in table 4.1 and the input variables in table 4.3. Just like in the tHq jet assignment, no quantities depending on the reconstructed Higgs boson are used as input variables in the tHW JA-BDT. The shapes of the tHW JA-BDT output for correct and wrong assignments can be seen in figure 4.7. No overtraining is observed. The agreement of simulation and data is presented in figure 4.8.

Table 4.3.: Input variables of the tHW JA-BDT. The variables are ranked by their importance in the training performed in sample A. Variations of the ranking in sample A and B begin with the variable ranked 3th in the training.

Variable	Description
$\log m(\mathrm{t})$	Invariant mass of the reconstructed top quark
$\log m(\mathrm{W_b})$	Invariant mass of the W boson from the top quark production vertex
$p_\mathrm{T}(\mathrm{t})$	Transverse momentum of the reconstructed top quark
$p_\mathrm{T}(\mathrm{W_b})$	Transverse momentum of the of the reconstructed W boson from the top quark production vertex
$\cos\theta(\mathrm{b_t}, l)$	Cosine of the angle between the jet assigned to the bottom quark from the top quark decay and the lepton
$\Delta R(\mathrm{b_t}, \mathrm{W_{lep}})$	ΔR between the jet assigned to thebottom quark from the top quark decay and the leptonically decaying W boson
$\lvert\eta(\mathrm{W_b})\rvert$	Absolute pseudorapidity of the reconstructed W boson from the top quark production vertex
$\lvert\eta(\mathrm{t})\rvert$	Absolute pseudorapidity of the reconstructed top quark
$\lvert\eta(\mathrm{b_t})\rvert$	Absolute pseudorapidity of the jet assigned to the bottom quark from the top quark decay
$\lvert\eta(\mathrm{t}) - \eta(\mathrm{W_b})\rvert$	Absolute difference of the pseudorapidity of the reconstructed top quark and the reconstructed W boson from the top quark production vertex
$\mathrm{CSV(b_t)}$	Output of the b tagging discriminant for the jet assigned to the bottom quark from the top quark decay



(a) (b)

Figure 4.7.: Distribution of the tHW JA-BDT output. The figure shows the distribution of the tHW JA-BDT output trained in sample A (a) and trained in sample B (b) for wrong (blue) and correct (red) jet assignments. The distributions of the training samples (dashed) and test samples (markers) look similar. No overtraining is observed.

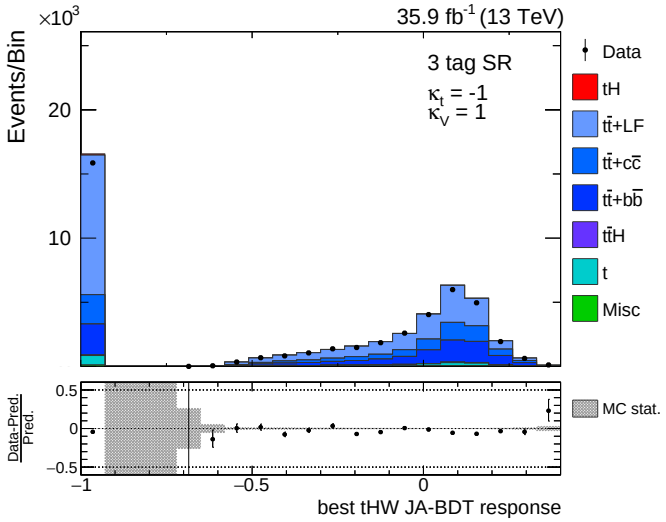


Figure 4.8.: Distribution of the best tHW JA-BDT response. The distribution of the best tHW JA-BDT response for the ITC scenario is shown for the 3 tag SR. Agreement of simulation and data is observed.

The diagram shows the distribution of the best tHW JA-BDT response for the ITC case in the 3 tag SR.

In the reference analysis, a tHW JA-BDT trained with the configuration given in table A.1 of appendix A and the input variables listed in table A.3 of appendix A are used.

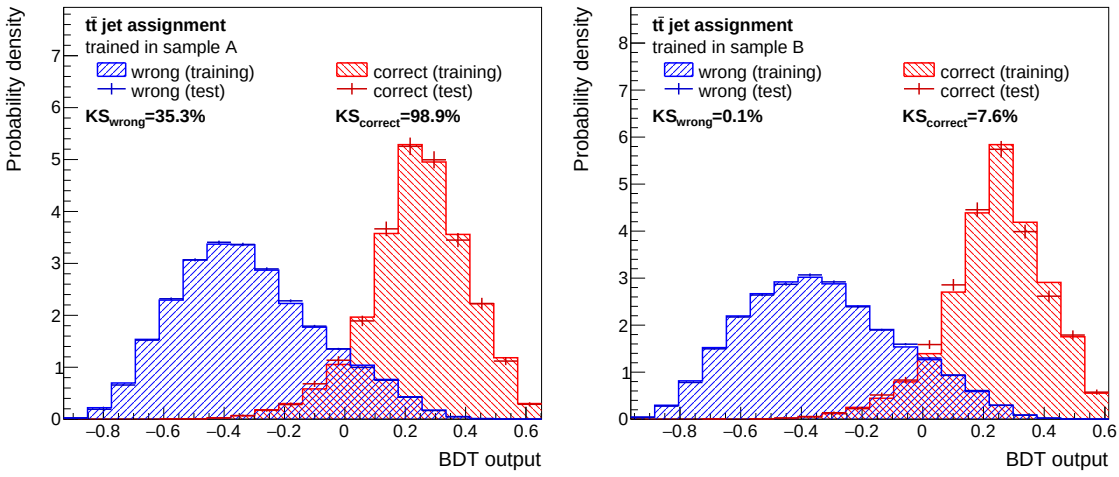
4.4.3. Jet Assignment under the $t\bar{t}$ Hypothesis

The two bottom quarks from the top quark decays and the two light-flavoured quarks from the hadronically decaying top quark are assigned to the reconstructed jets. The assignment procedure is analogous to the tHq and tHW jet assignment.

The configuration of the $t\bar{t}$ JA-BDT is given in table 4.1. The input variables of the $t\bar{t}$ JA-BDT are given in table 4.4. Figure 4.9 shows the distribution of the $t\bar{t}$ JA-BDT output for the training sample and an independent test sample. No overtraining is observed. The distribution of the best $t\bar{t}$ JA-BDT response can be found in figure 4.10. The distribution of the simulation accords with the distribution of the data. Since no Higgs boson has to be assigned to the jets in the $t\bar{t}$ process, this JA-BDT is the same as the one used in the reference analysis.

Table 4.4.: Input variables of the $t\bar{t}$ JA-BDT. The variables are ranked by their importance in the training performed in sample A. Variations of the ranking in sample A and B begin with the variable ranked 6th in the training.

Variable	Description
$\log(m(\mathrm{t_{had}}) - m(\mathrm{W_{had}}))$	Difference between the invariant mass of the reconstructed hadronically decaying top quark and the reconstructed hadronically decaying W boson
$\log m(\mathrm{W_{had}})$	Invariant mass of the two jets assigned to the reconstructed W boson from the hadronically decaying top quark
$\log m(\mathrm{t_{lep}})$	Invariant mass of the reconstructed leptonically decaying top quark
$\log p_\mathrm{T}(\mathrm{t_{had}})$	Transverse momentum of the reconstructed hadronically decaying top quark
CSV($\mathrm{W_{had}}$ jet 2)	Output of the b tagging discriminant for the jet assigned to the hadronically decaying W boson with the second highest transverse momentum
CSV($\mathrm{W_{had}}$ jet 1)	Output of the b tagging discriminant for the jet assigned to the hadronically decaying W boson with the highest transverse momentum
$\Delta R(\mathrm{W_{had}}$ jets)	ΔR between the two tight jets assigned to the hadronically decaying W boson
relative H_T	Fraction of the total transverse momenta that falls to the reconstructed hadronically and leptonically decaying top quarks
$\log p_\mathrm{T}(\mathrm{t_{lep}})$	Transverse momentum of the reconstructed leptonically decaying top quark



(a) (b)

Figure 4.9.: Distribution of the $t\bar{t}$ JA-BDT output. The figure shows the distribution of the $t\bar{t}$ JA-BDT output trained in sample A (a) and trained in sample B (b) for wrong (blue) and correct (red) jet assignments. The distributions of the training samples (dashed) and test samples (markers) look similar. No overtraining is observed.

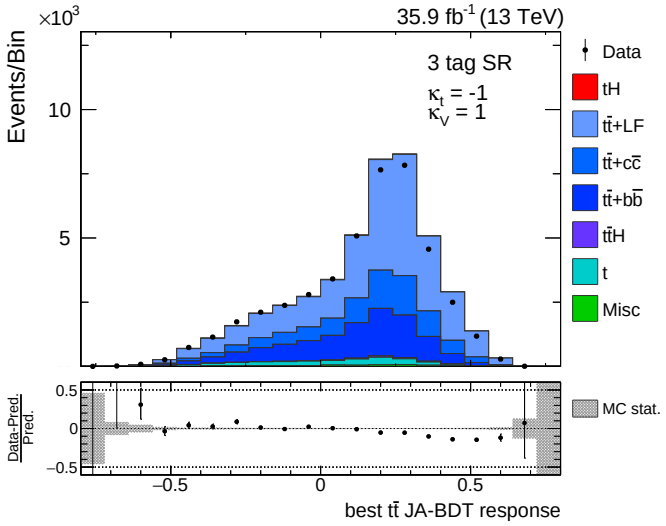


Figure 4.10.: Distribution of the best $t\bar{t}$ JA-BDT response. The distribution of the best $t\bar{t}$ JA-BDT response for the ITC scenario is shown for the 3 tag SR. Since the $t\bar{t}$ process is independent of the relative sign of κ_t and κ_V, the distribution is the same for all coupling scenarios. Agreement of simulation and data is observed.

4.4.4. Reducing Fluctuations in the Background Hypothesis

The choice of the correct and wrong jet-to-quark assignment is essential for the training of the JA-BDTs. In the previous subsections, the correct assignment was explained in detail. All assignments not defined as correct are referred to as wrong. Since there is not only one possible wrong assignment, they have to be treated differently than the correct assignments.

In previous analyses, this was done by taking randomly one wrong assignment for each event. This was then treated as if there were no other wrong assignments. Therefore, another jet-to-quark assignment was used as the wrong one in each new creation of background hypotheses. In sample A and sample B were also different wrong assignments taken for the training. The training of the BDTs depends on the choice of the wrong assignment which causes different performances in the application of every separately trained BDT.

An improved approach is to take all possible wrong assignments into account. For this purpose, the wrong assignments are weighted via

$$w_i = \frac{1}{N_{\mathrm{a}i}}\,. \tag{4.5}$$

The number of possible assignments $N_{\mathrm{a}i}$ for each event i depends on the number N_{j} of assignable jets (see equation (4.1) and (4.3)). The weights w_i ensure that all events are taken into account equally, no matter how many wrong assignments are possible. As the sum over all weights w_i per event adds up to one, the correct and wrong assignments are weighted the same.

In this thesis, the consideration of all wrong assignments with respective weights is implemented. As a result, the jet assignment is more stable, which means that the shapes of the BDT output from trainings belonging to different creations of hypotheses

look exactly the same. The shapes of the input variables and the BDT output are also more similar in sample A and B. The small difference can be accredited to statistical fluctuations. The sole disadvantage is the increase in computing time caused by much bigger samples including all wrong assignments and their respective weights.

All jet assignments presented in the subsections 4.4.1 to 4.4.3 already include this improvement.

4.5. Event Classification

After the jet assignment under the three hypotheses tHq, tHW, and $t\bar{t}$, the events are classified. The events belonging to the 3 tag SR and 4 tag SR are classified with a signal classification BDT (SC-BDT) according to their likeness to the tH signal process. Different global event variables as well as the variables depending on the jet assignments can be taken as input variables for the SC-BDT. The training procedure is the same as for the jet assignments described in section 4.4 and illustrated in figure 4.4. The tH processes are treated as signal processes resulting in a high BDT output value and the $t\bar{t}$ processes of all categories are taken as background processes (low BDT output). The impact of other background processes is too low to be taken into account in the signal classification. Analogue to the jet assignments, the training of the SC-BDT is performed for the ITC case. The distributions belonging to the other coupling scenarios are calculated using a reweighting procedure for the events. As a result, the shapes of the SC-BDT output look different for each coupling scenario.

In this thesis, variables depending on five different JA-BDTs can be simultaneously used: the variables from the three JA-BDTs of the reference analysis and from the tHq and tHW JA-BDT presented previously, which do not exploit Higgs-boson-dependent variables for the jet assignment. The aim is to improve the performance of the SC-BDT by increasing the area under the ROC curve. By trying different combinations of those variables as input variables of the SC-BDT, various ideas what could lead to an improvement are tested. In addition to the variables that were already used in previous analyses, the corresponding variables of the JA-BDTs independent of any Higgs-boson-dependent variable are added one after the other. The previous variables are also replaced successively by the corresponding variables of the JA-BDTs independent of any Higgs-boson-dependent variable. Each configuration of variables is first probed with a bigger number of trees. This normally leads to some overtraining and a greater area under the ROC curve. If the area under the ROC curve is not increased, the tested variable can be rejected. If it is, the number of trees and the other parameters can be optimised to avoid overtraining effects. In case the area under the ROC curve is still larger than that of the reference afterwards, this configuration can be used to check if this configuration leads to an improvement of the extracted limits. Since the area under the ROC curve is equal or smaller than that of the reference for every tested combination of input variables, the tests do not lead to any improvement.

To test the impact of the Higgs-boson-dependent variables in the jet assignment on the upper limits calculated within this analysis, the SC-BDT is trained with the same input variables and configuration as the reference. This configuration of the SC-BDT is stated in table 4.5. The variables used for the discrimination are given in table 4.6 and the shape

Table 4.5.: Configuration of the SC-BDT and the FC-BDT. The table contains the configuration of the SC-BDT and the FC-BDT. They are the same for the analysis described in this thesis and the reference. The percentages are based on the total number of events.

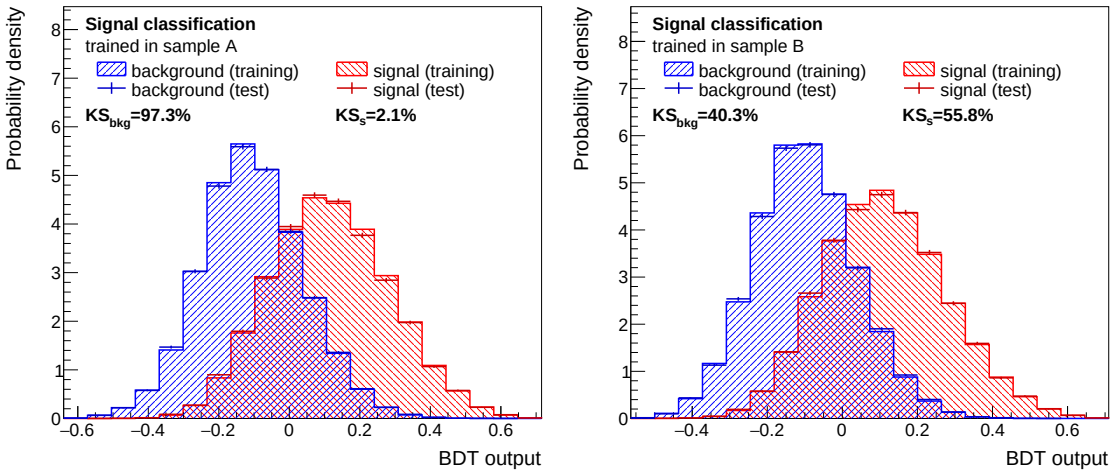
Parameter	SC-BDT	FC-BDT
`NTrees`	900	600
`MinNodeSize`	1 %	2 %
`MaxDepth`	4	2
`BoostType`	AdaBoost	AdaBoost
`AdaBoostBeta`	0.3	0.3
`nCuts`	15	16
`SeparationType`	GiniIndex	GiniIndex

of the SC-BDT output for sample A and B can be seen in figure 4.11. No overtraining is detected. The agreement of data and simulation for the SC-BDT response is shown in figure 4.12. The difference of both analyses results from different JA-BDTs. The reference analysis uses additional input variables depending on Higgs boson information in the jet assignment. The tHq and tHW JA-BDTs of this analysis are trained without using any variables depending on information of the reconstructed Higgs boson.

The events belonging to the 2 lep BR are classified with a flavour classification BDT (FC-BDT) according to their resemblance to the $t\bar{t} + \text{LF}$ background process. The FC-BDT is not changed for this analysis. The reference analysis and the presented analysis take the FC-BDT described in reference [10]. The used configuration can be found in table 4.5.

Table 4.6.: Input variables of the SC-BDT. The list is ordered by global event variables, tHq jet assignment variables, tHW jet assignment variables and $\mathrm{t\bar{t}}$ jet assignment variables. In each category, the variables are ranked by their importance in the training performed in sample A. Variations of the ranking in sample A and B begin with the top ranked variable. In the reference, the same input variables depending on other JA-BDTs are used.

Variable	Description
	Global event variables
aplanarity	Aplanarity [37] of the event
$\log m_3$	Invariant mass of the three jets that result in the highest p_T when combined
$q(l)$	Electric charge of the lepton
Fox-Wolfram #1	First Fox-Wolfram moment [38] of the event
	tHq jet assignment variables
tHq JA-BDT response	Output of the tHq JA-BDT
$\log p_\mathrm{T}(\mathrm{Higgs})$	p_T of the reconstructed Higgs boson candidate
$\log m(\mathrm{Higgs})$	Invariant mass of the reconstructed Higgs boson candidate
$\lvert\eta(\mathrm{light\,jet})\rvert$	Absolute pseudorapidity of the light forward jet
CSV(Higgs jet 2)	Output of the b tagging discriminant for the jet assigned to the Higgs boson with the second highest p_T
CSV(Higgs jet 1)	Output of the b tagging discriminant for the jet assigned to the Higgs boson candidate with the highest p_T
$\cos\theta(\mathrm{b_t}, l)$	Cosine of the angle between the jet assigned to the bottom quark from the top quark decay and the lepton
$\log p_\mathrm{T}(\mathrm{light\,jet})$	p_T of the light forward jet
$\cos\theta^*$	Cosine of the angle between the light forward jet and the lepton in the top quark frame
$\lvert\eta(\mathrm{t}) - \eta(\mathrm{Higgs})\rvert$	Absolute difference of the pseudorapidity of the reconstructed top quark and the reconstructed Higgs boson candidate
	tHW jet assignment variables
$\log\lvert p_\mathrm{T}(\mathrm{W_b})\rvert$	p_T of the reconstructed W boson from the top quark production vertex
tHW JA-BDT response	Output of the tHW JA-BDT
$\log\lvert m(\mathrm{W_b})\rvert$	Invariant mass of the W boson from the top quark production vertex
	$\mathrm{t\bar{t}}$ jet assignment variables
$\mathrm{t\bar{t}}$ JA-BDT response	Output of the $\mathrm{t\bar{t}}$ JA-BDT
CSV($\mathrm{W_{had}}$ jet 2)	Output of the b tagging discriminant for the jet assigned to the hadronically decaying W boson with the second highest p_T
CSV($\mathrm{W_{had}}$ jet 1)	Output of the b tagging discriminant for the jet assigned to the hadronically decaying W boson with the highest p_T
$\log m(\mathrm{t_{had}})$	Invariant mass of the reconstructed hadronically decaying top quark
$\Delta R(\mathrm{W_{had}\,jets})$	ΔR between the two light jets assigned to the hadronically decaying W boson



(a) (b)

Figure 4.11.: Distribution of the SC-BDT output. The figure shows the distribution of the SC-BDT output trained in sample A (a) and trained in sample B (b) for the background (blue) and the signal (red). The distributions of the training samples (dashed) and test samples (markers) look similar. No overtraining is observed.

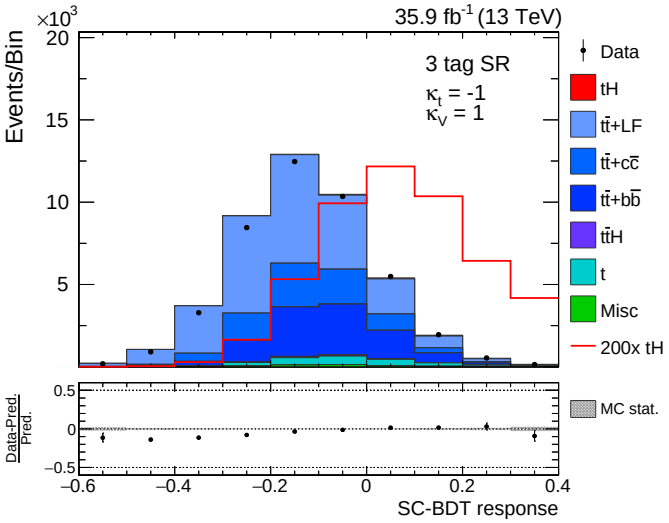


Figure 4.12.: Distribution of the SC-BDT response. The distribution of the SC-BDT response for the ITC scenario is shown for the 3 tag SR. Agreement of simulation and data is observed. The number of bins and the range of the x-axis correspond to them taken in the limit extraction described in chapter 4.7.

4.6. Uncertainty Treatment

In the analysis, different uncertainties are taken into account. In the final fit, which extracts the upper limit of the cross section times branching fraction, they are represented by nuisance parameters (see section 2.2). The uncertainties are of two types: shape and rate uncertainties. Shape uncertainties change only the distribution of the SC-BDT and FC-BDT output. Rate uncertainties change the total number of events. In general, uncertainties can be grouped in three categories: experimental, statistical and theoretical uncertainties. The uncertainties considered in this analysis are listed in the following for each group. It is specified in brackets whether they are shape or rate uncertainties.

Experimental Uncertainties Variations of the luminosity (rate), the jet energy scale (shape and rate), the jet energy resolution (shape and rate) and the contribution of unclustered particles to $p_{\mathrm{T}}^{\mathrm{miss}}$ (shape and rate) are taken into account. The uncertainty on the expected number of pileup interactions (shape and rate) and several uncertainties associated with the reweighting procedure (shape and rate) belong to this category, too. The uncertainties of the lepton efficiencies (shape and rate) are also taken into account.

Statistical Uncertainties The bin-by-bin uncertainties (shape) take into account that the number of events in each bin is affected by statistical variations caused by the limited size of analysed simulation samples.

Theoretical Uncertainties Further uncertainties are given by the choice of the Q^2 scale (shape and rate), the Q^2 scale of tW (rate) and the PDF set (rate). The effect of additional heavy-flavour jets (rate) and differences of the p_{T} spectra of the top quark (shape and rate) are also taken into account.

4.7. Exclusion Limits

For the different coupling scenarios specified by the ratio $\kappa_{\mathrm{t}}/\kappa_{\mathrm{V}}$, the upper limits on the cross section times branching fraction for the combined tH and $\mathrm{t\bar{t}H}$ process are determined by a simultaneous maximum likelihood fit of the SC-BDT output and the FC-BDT output. The $\mathrm{t\bar{t}H}$ process is taken in the fit as a signal process to increase the sensitivity to the absolute value of κ_{t}.

These upper limits at 95 % C.L. are depicted for the reference analysis (green lines) and presented analysis (yellow lines) in figure 4.13. The expected upper limits of the reference analysis are smaller than those of the presented analysis. Since the expected upper limits indicate the sensitivity of the analysis, the use of Higgs-boson-dependent variables in the jet assignment is an improvement for the investigation of different coupling scenarios. The observed limits lie within the bands corresponding to one standard deviation. The observations do not deviate significantly from the expectations.

The average improvement as a result of using reconstructed Higgs-boson-dependent variables in the jet assignment is 28.3 % (10.0 %) for the observed (expected) limit. The maximum (minimum) improvement of the observed limits is 31.4 % (20.1 %) for the coupling scenarios represented by $\kappa_{\mathrm{t}}/\kappa_{\mathrm{V}} = -4.00$ ($\kappa_{\mathrm{t}}/\kappa_{\mathrm{V}} = 1.25$). The maximum (minimum)

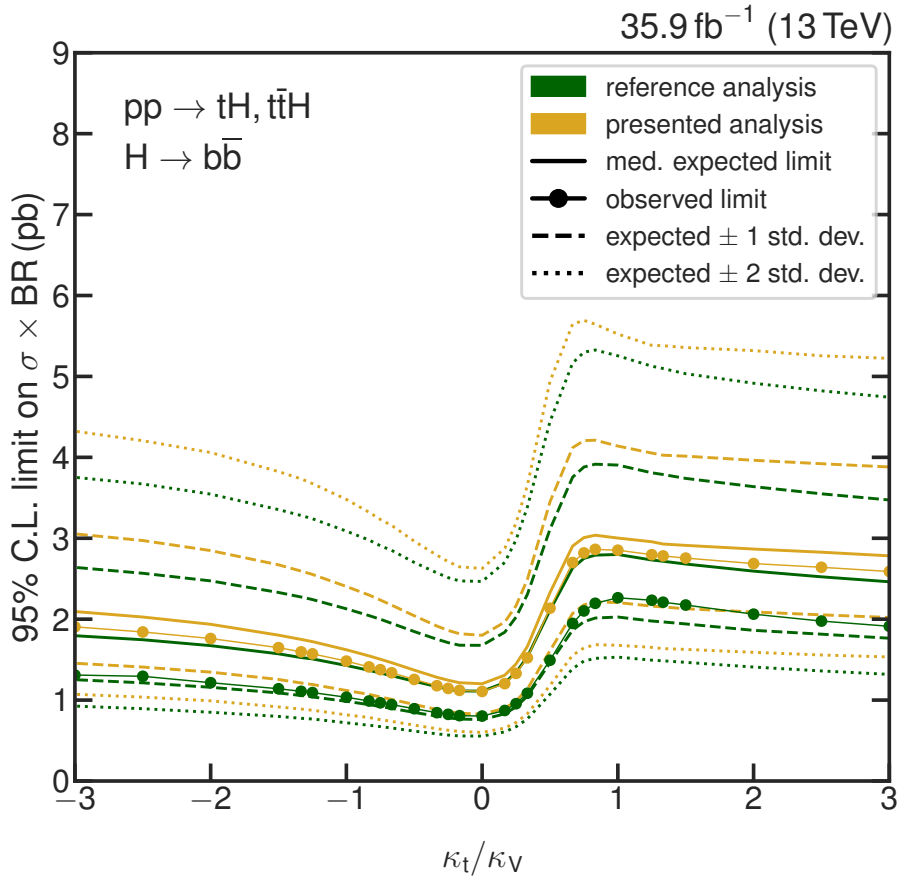


Figure 4.13.: Improvement of the limits accomplished with the use of Higgs-boson-dependent variables in the jet assignment. The impact of the use of Higgs-boson-dependent variables in the JA-BDTs on the expected and observed limits on the cross section times branching fraction of the combined tH and tt̄H process is illustrated for different coupling scenarios given by κ_t/κ_V. The expected (continuous lines) and observed (dots) limits as well as the one (dashed lines) and two (dotted lines) standard deviations from the expected limits are indicated for the reference analysis (green) and the presented analysis (yellow). No significant deviations from the expectation are observed.

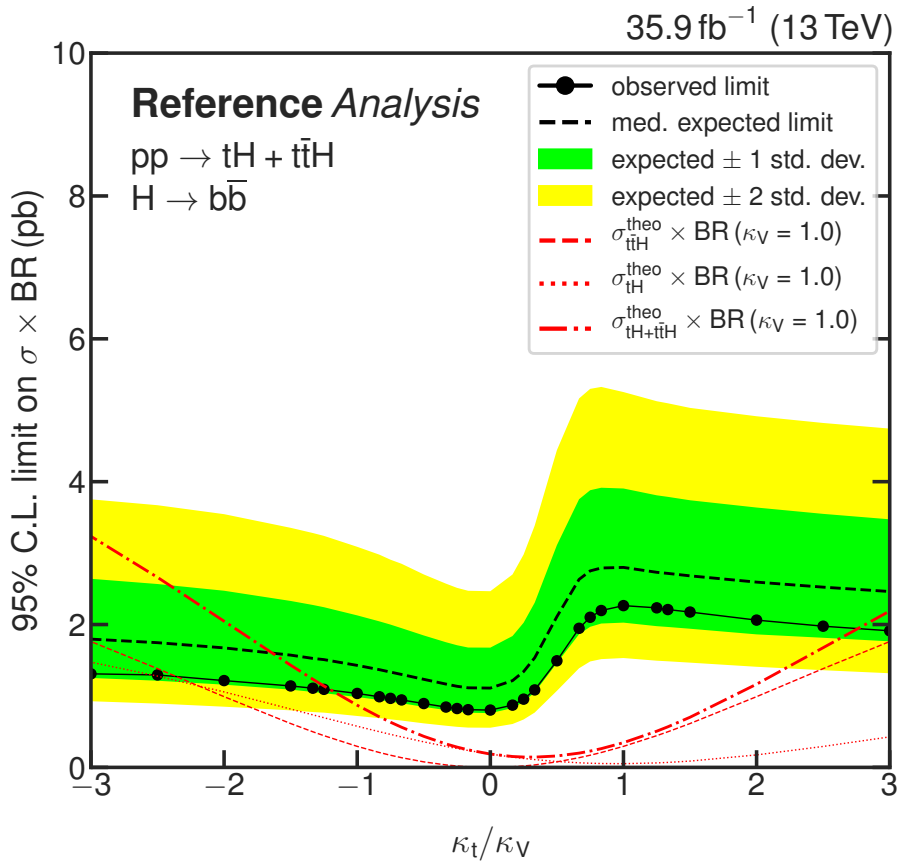


Figure 4.14.: Upper limits on the cross section times branching fraction for the tH + tt̄H process. The upper limits on the cross sections times branching fraction for the combined tH and tt̄H process of the reference analysis are shown for different ratios κ_t/κ_V of the coupling strength modifiers of the Higgs boson to the top quark κ_t and the W boson κ_V. The theoretical predictions for $\kappa_V = 1$ are shown for different processes with red lines. The green (yellow) band indicates one (two) standard deviations from the expectation. No significant deviations are observed.

improvement of the expected limits is given by 14.6 % (6.8 %) in the case of $\kappa_t/\kappa_V = -6.00$ ($\kappa_t/\kappa_V = 1.00$). The improvements of the expected and observed limits of each scenario are given in appendix B in table B.1 and B.2.

Figure 4.14 shows the reference analysis and the theoretical predicted cross section times branching fraction for the tt̄H, tH and combined tH + tt̄H process. A specific coupling scenario can be excluded if the observed cross section times branching fraction is smaller than predicted by the theory. With the reference analysis, the coupling scenarios given by $-6.00 \leq \kappa_t/\kappa_V \leq -1.25$ and $3.00 \leq \kappa_t/\kappa_V \leq 6.00$ can be excluded.

5. Summary and Outlook

The analysis presented in this thesis studies different coupling scenarios of the Higgs boson to top quarks and vector bosons. The coupling strength of the Higgs boson to top quarks g_V and the Yukawa coupling strength of the Higgs boson to fermions y_f are investigated. Deviations from the SM can be written as $\kappa_t = y_t/y_t^{SM}$ and $\kappa_V = g_V/g_V^{SM}$. The different coupling scenarios are thus given by the fraction κ_t/κ_V. The SM corresponds to $\kappa_t = \kappa_V = 1$ and the inverted top coupling (ITC) case is given by $-\kappa_t = \kappa_V = 1$. For the investigation of the different coupling scenarios, the combined tH and $t\bar{t}H$ production is treated as the signal process. The jet-to-quark assignment under different hypotheses is realised by BDTs called jet assignment BDTs. Within this thesis, the impact of input variables depending on Higgs boson information in the jet assignment on the sensitivity of the analysis is studied.

Compared to the analysis that uses no variables depending on the Higgs boson information in the jet assignment, the observed (expected) upper limits at 95 % C.L. of the reference analysis on the cross section times branching fraction for the combined tH and $t\bar{t}H$ process are 28.3 % (10.0 %) lower on average. The observed (expected) upper limits of the reference analysis in which Higgs-boson-dependent variables are used in the jet assignment are given by 2.26 pb (2.80 pb), which corresponds to $6.53 \times \sigma_{SM}$ ($8.09 \times \sigma_{SM}$) for the SM, and 1.04 pb (1.43 pb), which corresponds to $1.19 \times \sigma_{ITC}$ ($1.64 \times \sigma_{ITC}$) for the ITC scenario. Thus, the observed limit is 20.6 % (30.2 %) better for the SM (ITC scenario). The expected limit is 6.8 % (12.0 %) lower for the SM (ITC scenario). With the reference analysis, the coupling scenarios given by $\kappa_t/\kappa_V \leq -1.25$ and $3.00 \leq \kappa_t/\kappa_V$ can be excluded. The ITC case can also nearly be excluded.

Thus, this thesis shows that using Higgs-boson-dependent variables in the jet assignment, as done in the reference analysis and earlier published versions of the analysis [39], leads to improved results.

The improvement realised within this thesis is the reduction of fluctuations in the background hypothesis. In the training of the jet assignment BDTs, all possible wrong jet assignments are now taken into account instead of only one randomly chosen wrong jet assignment. This improves the stability of the jet assignment and thereby of the whole analysis.

Taking a combination of variables from the jet assignments, which gets Higgs-boson-dependent variables as input variables and the jet assignment which uses only variables independent of Higgs boson information as input variables of the event classification BDT is also tested. So far, this is not able to improve the upper limits on the cross section times branching fraction. However, in the future in the case of larger simulated samples additional variables with separation power are necessary. For this, the variables from the jet assignment which uses only variables independent of Higgs boson information could be taken.

A. Jet Assignment of the Reference Analysis

Table A.1.: Configuration of the JA-BDTs of the reference. The table contains the configuration of the JA-BDTs of the reference used for the jet assignment under the tHq, tHW and $t\bar{t}$ hypothesis.

Parameter	tHq JA-BDT	tHW JA-BDT	$t\bar{t}$ JA-BDT
`NTrees`	700	800	400
`MinNodeSize`	1 %	1 %	1 %
`MaxDepth`	4	5	3
`BoostType`	AdaBoost	AdaBoost	AdaBoost
`AdaBoostBeta`	0.3	0.3	0.3
`nCuts`	20	20	20
`SeparationType`	GiniIndex	GiniIndex	GiniIndex

Table A.2.: Input variables of the tHq JA-BDT of the reference. The variables are ranked by their importance in the training performed in sample A. Variations of the ranking in sample A and B begin with the variable ranked 5$^{\text{th}}$ in the training.

Variable	Description
$\log m(\mathrm{H})$	Invariant mass of the reconstructed Higgs boson
$\Delta R(\mathrm{b_t}, l)$	ΔR between the jet assigned to the bottom quark from the top quark decay and the lepton
$p_{\mathrm{T}}(\mathrm{t})$	p_{T} of the reconstructed top quark
$\log m(\mathrm{t})$	Invariant mass of the reconstructed top quark
$\cos\theta(\mathrm{b_t}, l)$	Cosine of the angle between the jet assigned to the bottom quark from the top quark decay and the lepton
relative H_{T}	Fraction of the total transverse momenta that falls to the b jet from the top quark, Higgs jet and light forward jet
$p_{\mathrm{T}}(\mathrm{H})$	p_{T} of the reconstructed Higgs boson
$\Delta R(\mathrm{Higgs\,jets})$	ΔR between the two jets from the Higgs boson decay
$\lvert\eta(\mathrm{b_t})\rvert$	Absolute pseudorapidity of the jet assigned to the bottom quark from the top quark decay
$p_{\mathrm{T}}(\mathrm{light\,jet})$	Transverse momentum of the light forward jet
$\Delta R(\mathrm{t, H})$	ΔR between the reconstructed top quark and reconstructed Higgs boson
$\lvert\eta(\mathrm{light\,jet}) - \eta(\mathrm{b_t})\rvert$	Absolute difference of the pseudorapidity of the light forward jet and the bottom quark from the top quark decay
CSV($\mathrm{b_t}$)	Output of the b tagging discriminant for the jet assigned to the bottom quark from the top quark decay
$\log\min(p_{\mathrm{T}}(\mathrm{H\,jets}))$	Minimum of the transverse momenta of the two jets assigned to the Higgs boson decay products
$\lvert\eta(\mathrm{t})\rvert$	Absolute pseudorapidity of the reconstructed top quark
CSV(Higgs jet 1)	Output of the b tagging discriminant for the jet assigned to the Higgs boson with the highest p_{T}
$\lvert\eta(\mathrm{light\,jet})\rvert$	Absolute pseudorapidity of the light forward jet
CSV(Higgs jet 2)	Output of the b tagging discriminant for the jet assigned to the Higgs boson with the second highest p_{T}
$\lvert\eta(\mathrm{t}) - \eta(\mathrm{H})\rvert$	Absolute difference of the pseudorapidity of the reconstructed top quark and the reconstructed Higgs boson
$\Delta E(\mathrm{light\,jet, b_t})$	Jet energy difference of the light forward jet and the jet assigned to the bottom quark from the top quark decay
$\lvert\eta(\mathrm{H})\rvert$	Absolute pseudorapidity of the reconstructed Higgs boson

Table A.3.: Input variables of the tHW JA-BDT. The variables are ranked by their importance in the training performed in sample A. Variations of the ranking in sample A and B begin with the variable ranked 6$^{\text{th}}$ in the training.

Variable	Description
$\log m(\mathrm{H})$	Invariant mass of the reconstructed Higgs boson
$\log m(\mathrm{t})$	Invariant mass of the reconstructed top quark
$\log m(\mathrm{W_b})$	Invariant mass of the W boson from the top quark production vertex
$\Delta R(\text{Higgs jets})$	ΔR between the two jets from the Higgs boson decay
$\Delta R(\mathrm{b_t}, \mathrm{W_{lep}})$	ΔR between the jet assigned to the bottom quark from the top quark decay and the leptonically decaying W boson
$\cos\theta(\mathrm{b_t}, l)$	Cosine of the angle between the jet assigned to the bottom quark from the top quark decay and the lepton
$p_{\mathrm{T}}(\mathrm{W_b})$	Transverse momentum of the reconstructed W boson from the top quark production vertex
$p_{\mathrm{T}}(\mathrm{t})$	Transverse momentum of the reconstructed top quark
relative H_{T}	Fraction of the total transverse momenta that falls to the b jet from the top quark, Higgs jet and light forward jet
$p_{\mathrm{T}}(\mathrm{H})$	Transverse momentum of the reconstructed Higgs boson
$\lvert\eta(\mathrm{b_t})\rvert$	Absolute pseudorapidity of the jet assigned to the bottom quark from the top quark decay
$\lvert\eta(\mathrm{W_b})\rvert$	Absolute pseudorapidity of the reconstructed W boson from the top quark production vertex
$\lvert\eta(\mathrm{t})\rvert$	Absolute pseudorapidity of the reconstructed top quark
$\lvert\eta(\mathrm{t}) - \eta(\mathrm{H})\rvert$	Absolute difference of the pseudorapidity of the reconstructed top quark and the reconstructed Higgs boson
CSV(Higgs jet 1)	Output of the b tagging discriminant for the jet assigned to the Higgs boson with the highest transverse momentum
CSV($\mathrm{b_t}$)	Output of the b tagging discriminant for the jet assigned to the bottom quark from the top quark decay
CSV(Higgs jet 2)	Output of the b tagging discriminant for the jet assigned to the Higgs boson with the second highest transverse momentum
$\lvert\eta(\mathrm{t}) - \eta(\mathrm{W_b})\rvert$	Absolute difference of the pseudorapidity of the reconstructed top quark and the reconstructed W boson from the top quark production vertex

B. Improvement Accomplished with Higgs-Boson-Dependent Variables in the Jet Assignment

Table B.1.: Improvement of the expected limits via Higgs-boson-dependent variables in the jet assignment. The expected limits on the cross section times branching fraction ($\sigma \times$ BR) of combined $tH + t\bar{t}H$ process with different coupling scenarios represented by the ratio κ_t/κ_V are given for the presented analysis and the reference analysis. The improvement of the reference analysis caused by the use of Higgs-boson-dependent variables in the jet assignment is also listed. The expected limit is enhanced for all scenarios.

Scenario given by κ_t/κ_V	expected limit on $\sigma \times$ BR(pb)		improvement in %
	presented analysis	reference analysis	
−6.00	2.28	1.94	14.63
−4.00	2.18	1.87	14.37
−3.00	2.09	1.80	14.25
−2.50	2.03	1.75	13.85
−2.00	1.94	1.67	13.62
−1.50	1.80	1.57	12.80
−1.33	1.75	1.53	12.81
−1.25	1.72	1.51	12.40
−1.00	1.62	1.43	11.95
−0.83	1.55	1.37	11.53
−0.75	1.50	1.33	11.33
−0.67	1.46	1.30	11.04
−0.50	1.37	1.23	10.04
−0.33	1.28	1.17	9.01
−0.25	1.25	1.14	8.82
−0.17	1.21	1.11	8.29
0.00	1.20	1.11	7.64
0.17	1.31	1.21	7.10
0.25	1.45	1.34	7.28
0.33	1.66	1.53	7.69
0.50	2.33	2.11	9.32
0.67	2.90	2.63	9.35
0.75	3.01	2.75	8.54
0.83	3.04	2.79	8.10
1.00	3.00	2.80	6.76
1.25	2.96	2.73	7.63
1.33	2.93	2.72	7.27
1.50	2.91	2.68	7.91
2.00	2.87	2.59	9.52
2.50	2.83	2.52	10.76
3.00	2.78	2.46	11.52
4.00	2.73	2.38	12.72
6.00	2.66	2.30	13.65

Table B.2.: Improvement of the observed limits via Higgs-boson-dependent variables in the jet assignment. The observed limits on the cross section times branching fraction ($\sigma \times$ BR) of combined $tH + t\bar{t}H$ process with different coupling scenarios represented by the ratio κ_t/κ_V are given for the presented analysis and the reference analysis. The improvement of the reference analysis caused by the use of Higgs-boson-dependent variables in the jet assignment is also listed. The observed limit is enhanced for all scenarios.

Scenario given by κ_t/κ_V	**observed limit on $\sigma \times$ BR(pb)**		**improvement in %**
	presented analysis	**reference analysis**	
−6.00	2.07	1.43	31.19
−4.00	1.99	1.36	31.41
−3.00	1.91	1.31	31.27
−2.50	1.84	1.29	29.74
−2.00	1.76	1.22	31.00
−1.50	1.65	1.14	30.82
−1.33	1.57	1.09	30.45
−1.00	1.48	1.04	30.15
−0.83	1.41	0.99	29.87
−0.75	1.38	0.97	29.79
−0.67	1.34	0.94	29.42
−0.50	1.26	0.89	28.81
−0.33	1.18	0.84	28.29
−0.25	1.14	0.82	28.00
−0.17	1.12	0.81	27.89
0.00	1.11	0.80	27.50
0.17	1.20	0.87	27.58
0.25	1.33	0.96	28.17
0.33	1.52	1.08	28.73
0.50	2.14	1.49	30.13
0.67	2.70	1.95	27.87
0.75	2.82	2.10	25.43
0.83	2.86	2.20	23.29
1.00	2.85	2.26	20.60
1.25	2.80	2.23	20.14
1.33	2.78	2.21	20.57
1.50	2.75	2.17	21.05
2.00	2.69	2.06	23.27
2.50	2.64	1.98	25.17
3.00	2.59	1.91	26.15
4.00	2.53	1.83	27.67
6.00	2.45	1.74	28.95

Bibliography

[1] B. Povh, K. Rith, C. Scholz, F. Zetsche, and W. Rodejohann, "Teilchen und Kerne: Eine Einführung in die physikalischen Konzepte", Springer-Lehrbuch, Springer Spektrum, Berlin, Heidelberg, 2014.

[2] "Standard Model of Elementary Particles", 2006. https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg, last accessed on 2019-04-14.

[3] M. Tanabashi *et al.*, "Review of Particle Physics", *Physical Review D*, vol. 98, 2018.

[4] N. Faltermann, "Single Top Quark Production at 13 TeV with the CMS Experiment: from Rediscovery to Search for Rare Channels and Determination of Higgs Boson Couplings". PhD thesis, Karlsruhe Institute of Technology (KIT), ETP-KA/2018-10, 2018.

[5] The CMS Collaboration, "Combined measurements of the Higgs boson's couplings at $\sqrt{s} = 13$ TeV", CMS-PAS-HIG-17-031, CERN, Geneva, 2018.

[6] G. Bohm and G. Zech, "Introduction to statistics and data analysis for physicists", Verl. Dt. Elektronen-Synchrotron, Hamburg, 2010.

[7] V. Blobel and E. Lohrmann, "Statistische und numerische Methoden der Datenanalyse", Teubner Studienbücher, Teubner Stuttgart, Leipzig, 1998.

[8] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss, "TMVA: Toolkit for Multivariate Data Analysis", *Proceeding of Science*, vol. ACAT, p. 040, 2007.

[9] R. Brun and F. Rademakers, "ROOT: An object oriented data analysis framework", *Nuclear Instruments and Methods*, vol. A389, pp. 81–86, 1997.

[10] D. Bühler, "Improvement of the b-Jet Energy Reconstruction via Multivariate Regression for the Investigation of Higgs Boson Couplings to Top Quarks and Vector Bosons", Master's thesis, Karlsruhe Institute of Technology (KIT), ETP-KA/2019-03, 2019.

[11] 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*, vol. 55, no. 1, pp. 119 – 139, 1997.

[12] "Overfitting", 2008. https://commons.wikimedia.org/wiki/File:Overfitting.svg, last accessed on 2019-04-01.

[13] T. Junk, "Confidence Level Computation for Combining Searches with Small Statistics", *Nuclear Instruments and Methods in Physics A*, vol. 434, pp. 435–443, 1999.

[14] The CMS Collaboration, "Documentation of the COMBINE package.", 2018. `https://cms-hcomb.gitbooks.io/combine/`, last accessed on 2019-04-02.

[15] The CMS Collaboration, "Welcome to the RooStats Wiki" `https://twiki.cern.ch/twiki/bin/view/RooStats/WebHome`, last accessed on 2019-04-01.

[16] CERN - Education, Communications and Outreach Group, "LHC Guide", Mar 2017. `http://cds.cern.ch/record/2255762`, last accessed on 2019-03-28.

[17] CERN, "The Large Hadron Collider", 2014. `http://home.cern/topics/large-hadron-collider`, last accessed on 2019-03-28.

[18] The ALICE Collaboration, "The ALICE experiment at the CERN LHC", *Journal of Instrumentation*, vol. 3, p. S08002, 2008.

[19] The LHCb Collaboration, "The LHCb Detector at the LHC", *Journal of Instrumentation*, vol. 3, p. S08005, 2008.

[20] The ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *Journal of Instrumentation*, vol. 3, p. S08003, 2008.

[21] S. Chatrchyan *et al.*, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Physics Letters B*, vol. 716, pp. 30–61, 2012.

[22] G. Aad *et al.*, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Physics Letters B*, vol. 716, pp. 1–29, 2012.

[23] P. Lujan, "CMS Luminosity - Public Results", 2018. `https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults`, last accessed on 2019-04-03.

[24] The CMS Collaboration, "The CMS Experiment at the CERN LHC", *Journal of Instrumentation*, vol. 3, p. S08004, 2008.

[25] CERN, "CMS: The Compact Muon Solenoid", 2012. `https://cds.cern.ch/record/1997263`, last acessed on 2019-04-03.

[26] D. Haitz, "Precision Studies of Proton Structure and Jet Energy Scale with the CMS Detector at the LHC". PhD thesis, Karlsruhe Institute of Technology (KIT), IEKP-KA/2016-04, 2016.

[27] S. R. Davis, "Interactive Slice of the CMS detector", 2016. `https://cds.cern.ch/record/2205172`, last accessed on 2019-04-01.

[28] The CMS Collaboration, "Precise mapping of the magnetic field in the CMS barrel yoke using cosmic rays", *Journal of Instrumentation*, vol. 5, p. T03021, 2010.

[29] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, "MadGraph 5 : Going Beyond", *Journal of High Energy Physics*, vol. 06, p. 128, 2011.

[30] S. Frixione and B. R. Webber, "Matching nlo qcd computations and parton shower simulations", *Journal of High Energy Physics*, vol. 2002, no. 06, p. 029, 2002.

[31] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method", *Journal of High Energy Physics*, vol. 11, p. 070, 2007.

[32] T. Sjöstrand *et al.*, "An Introduction to PYTHIA 8.2", *Computer Physics Communications*, vol. 191, pp. 159–177, 2015.

[33] S. Agostinelli *et al.*, "G4 - a simulation toolkit", *Nuclear Instruments and Methods*, vol. 506, pp. 250–303, 07 2003.

[34] S. Chatrchyan *et al.*, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *Journal of Instrumentation*, vol. 9, p. P10009, 2014.

[35] The CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector", *Journal of Instrumentation*, vol. 12, p. P10003, 2017.

[36] The CMS Collaboration, "Identification of b quark jets at the CMS Experiment in the LHC Run 2", CMS-PAS-BTV-15-001, CERN, Geneva, 2016.

[37] V. D. Barger, J. Ohnemus, and R. J. N. Phillips, "Event shape criteria for single lepton top signals", *Physical Review*, vol. D48, pp. R3953–R3956, 1993.

[38] G. C. Fox and S. Wolfram, "Observables for the Analysis of Event Shapes in e^+ e^- Annihilation and Other Processes", *Physical Review Letters*, vol. 41, p. 1581, 1978.

[39] The CMS Collaboration, "Search for associated production of a Higgs boson and a single top quark in proton-proton collisions at $\sqrt{s} = 13$ TeV", CMS-HIG-18-009, CERN, Geneva, 2018.

Danksagung

Bei Prof. Dr. Thomas Müller möchte ich mich für die Aufnahme in seine Arbeitsgruppe bedanken, wodurch er mir diese Bachelorarbeit ermöglichte. Dr. Thorsten Chwalek möchte ich für die Übernahme des Korreferats danken.

Bei Kevin Flöh bedanke ich mich für die wunderbare Betreuung und bei Darius Bühler dafür, dass er mir den Einstieg in die Teilchenphysik erleichterte.

Auch allen Korrekturlesenden sei an dieser Stelle herzlich gedankt: Dr. Thorsten Chwalek, Dr. Nils Faltermann, Kevin Flöh, Darius Bühler und Tabea Friedrich.

Bei der Mittagessen-Gruppe, insbesondere Darius Bühler, David Seith, Denise Müller, Ellen Sarauer, Johann Rauser, Kevin Flöh, Max Neukum, Neeraj Amin, Dr. Nils Faltermann, Tabea Friedrich und Dr. Thorsten Chwalek, bedanke ich mich für die lustigen Gespräche und dafür, dass sie immer geduldig auf mich warteten.

Ein besonderer Dank gilt Tabea Friedrich, die mich während des gesamten Bachelorstudiums als Freundin und Arbeitskollegin mit ihrer Gabe, alles bis ins kleinste Detail verstehen zu wollen, begleitete.

Für Rückhalt und Unterstützung möchte ich meiner Familie und meinen Freunden danken.