TECHNISCHE
UNIVERSITÄT
WIEN
Vienna University of Technology

DIPLOMARBEIT

Potential of the CMS $H \to \tau^+\tau^-$ analysis using boosted decision trees

AUSGEFÜHRT AM

Atominstitut Wien
in Verbindung mit dem Institut für Hochenergiephysik (HEPHY)
der Österreichischen Akademie der Wissenschaften (ÖAW)

UNTER DER ANLEITUNG VON

Privatdoz. Dr.rer.nat. Manfred Jeitler
Dr. Martin Flechl

EINGEREICHT AN

der Technischen Universität Wien
Fakultät für Physik

VON

Verena Schamböck
Seyringer Straße 33
2211 Pillichsdorf, Österreich

Ort, Datum

Unterschrift (Student)

Abstract

In 2012, a new boson with a mass of approximately 125 GeV was discovered at the Large Hadron Collider (LHC) at CERN by the ATLAS and CMS Collaboration. Its properties were found to have good agreement with the sought-after Higgs boson described in the Standard Model of particle physics. Until then, the Higgs particle was the last particle of the Standard Model that was not experimentally verified. One important step in confirming the boson as the Standard Model Higgs boson is the measurement of its coupling to τ leptons with high precision. The most promising decay channel to reach discovery significance is the $H \to \tau^+\tau^-$ decay. To enhance the performance of the event classification of data taken at LHC at a center-of-mass energy of 13 TeV, boosted decision trees (BDT) are studied as potential signal extraction method. BDT is a machine learning technique, which was optimized and trained by the use of Monte Carlo samples. As a benchmark, the BDT performance is compared to other signal-versus-background classifier like a cut-based method, which is the standard technique used in high energy physics, and to other machine learning methods. The BDT analysis performs significantly better than the other approaches and therefore has high potential for future data analysis.

Kurzzusammenfassung

Im Jahr 2012 wurde am Large Hadron Collider (LHC) am CERN von der CMS und ATLAS Collaboration ein neues Boson mit einer ungefähren Masse von 125 GeV entdeckt. Seine Eigenschaften stimmen gut mit jenen des gesuchten Higgs Bosons des Standard Modells der Teilchenphysik überein. Bis dahin war das Higgs Boson das letzte Teilchen des Standard Modells, das noch nicht experimentell beobachtet worden war. Ein wichtiger Schritt um die Übereinstimmung des Bosons mit dem Higgs Boson des Standard Modells zu bestätigen, ist die Messung der Kopplung zu τ Leptonen mit hoher Präzision. Der vielversprechendste Zerfallskanal, um die notwendige Signifikanz für eine Entdeckung zu erreichen, ist $H \to \tau^+\tau^-$. Um die Leistung der Eventklassifizierung von Datensätzen, die am LHC bei einer Schwerpunktsenergie von 13 TeV detektiert werden, zu erhöhen, werden Boosted Decision Trees (BDT) als potentielle Signalextraktionsmethode untersucht. BDT ist eine selbstlernende Methode, die durch die Verwendung von Monte Carlo Datensätzen optimiert und trainiert wird. Die BDTs werden mit den Ergebnissen anderer Methoden, die Signalereignisse von Hintergrundereignissen trennen, verglichen. Als Vergleich werden eine schnittbasierte Methode, die der Standardmethode in der Hochenergiephysik entspricht, und zwei selbstlernende Methoden herangezogen. Die BDTs weisen deutlich bessere Ergebnisse als die anderen Methoden auf und zeigen somit hohes Potential für die zukünftige Datenanalyse.

Table of Contents

1 Introduction

In the Standard Model of particle physics (SM) the Higgs mechanism plays an important role as it provides an explanation of the mass generation for the $W^\pm$ and Z gauge bosons and fermions [1]. It was predicted in 1964, implying the existence of a massive scalar boson, the Higgs boson [2, 3, 4]. The discovery of the $W^\pm$ and Z gauge bosons in 1983 [5, 6] gave strong indication for the Higgs mechanism. In 2012, a boson with a mass of 125 GeV was discovered at CERN by the ATLAS and CMS Collaborations. Its properties show good agreement with those of the SM Higgs boson [7, 8, 9]. The masses of fermions are generated by Yukawa couplings between the fermionic fields and the Higgs field [10]. The confirmation of this coupling is essential for the identification of the discovered boson as the SM Higgs boson. The Higgs boson decaying into a tau-antitau pair ($H \to \tau^+\tau^-$) is the most promising decay mode for such a measurement. Its expected event rate at the Large Hadron Collider (LHC) is high compared to most other fermionic decay channels and it has a relatively small background contribution with respect to decays to $b\bar{b}$ [11].

The purpose of this thesis is the investigation of the performance of Boosted Decision Trees (BDT) for the $H \to \tau\tau$ signal extraction. The analyses are performed for a center-of-mass energy of 8 TeV and 13 TeV with the focus lying on the 13 TeV analysis. The BDTs are trained on Monte Carlo (MC) samples simulating the different signal and background processes that are included in the analyses. Furthermore, the BDT performance is compared to a cut-based approach. The BDT analysis is conducted by the use of the Toolkit for Multivariate Analysis (TMVA), which is a ROOT-integrated environment.

The thesis starts out with an introduction to the SM focusing on spontaneous symmetry breaking and the Higgs boson (Chap. 2). Then, a brief overview on the LHC at CERN is given with an emphasis on the CMS detector and its subsystems (Chap. 3). Chapter 4 introduces the cut-based and multivariate analysis (MVA) approaches, gives an overview on the TMVA framework and describes the methods of boosted decision trees in more detail. In chapter 5 two possibilities for the performance quantification are explained. The MC samples, used for training and evaluating the BDT algorithm, are discussed in chapter 6. The 8 TeV MC samples are normalized to an luminosity of $\mathcal{L} = 19.7\,\text{fb}^{-1}$ and the 13 TeV to $\mathcal{L} = 10\,\text{fb}^{-1}$. The analysis for 8 TeV is performed for the $\mu\tau_h$ channel under consideration of the Drell-Yan background and discussed in chapter 7. The BDT optimization methods are also introduced in this chapter. In chapter 8 the different analyses performed for 13 TeV are explained and their results are compared. For 13 TeV not only the $\mu\tau_h$ channel, but also $e\tau_h$ is studied and a more com-

plete background treatment is implemented by including the $t\bar{t}$ and W+jets background. Thereafter, the performance of the 13 TeV analyses is quantified by the CMS statistics framework, which provides the possibility of the implementation of nuisance parameters (Chap. 9). All results are summarized in chapter 10, which also provides an outlook.

2 The Standard Model of particle physics

The SM is a theory describing all currently known elementary particles and their interactions [1]. These interactions include the strong force, the weak force and the electromagnetic force which are described by a gauge theory. As the SM is formulated as a local gauge theory, the forces are mediated via gauge bosons. In addition to the four gauge particles, 12 fermions, consisting of quarks and leptons, and the Higgs boson are included in the SM. Each fermion has an associated anti-fermion having opposite electric charge.

The SM is built on 5 fundamental assumptions, namely

- Homogeneity and isotropy of space

- Special relativity

- Gauge symmetry

- Spontaneous symmetry breaking

- CPT (Charge conjugation, Parity transformation, Time reversal) invariance

A big success of the SM is not only the description of the forces, but also the prediction of elementary particles and their properties. For example, the Higgs boson was predicted as a consequence of a spontaneous symmetry breaking in the SM in the 1960s [2, 3, 4]. In 2012, a boson was discovered at the LHC at CERN, which shows good agreement with the properties of the predicted Higgs boson [7, 8, 9].

Nevertheless, there are still a lot of open questions that can not be answered by the SM, only some of them will be mentioned here. Firstly, the fourth fundamental force, gravitation, is not part of the SM as it can not be described by gauge theory yet. Another open question is, if there exists a Grand Unified Theory (GUT) that unifies all fundamental forces. Moreover, what is dark energy and dark matter? Why are fermions arranged in 3 generations? Where are the missing CP (charge parity) symmetry breaking processes?

Many theories try to solve these problems, like supersymmetry (SUSY) or extra dimensions.

Nonetheless, only the SM is discussed in the present chapter, starting with a more detailed overview on its particles, then the fundamental forces are discussed and in the last chapter the Higgs mechanism is introduced by global and local symmetry breaking.

2.1 Particles of the Standard Model and fundamental forces

Two groups of particles can be distinguished, fermions and bosons. The fermions contain leptons and quarks, which have spin-1/2. The bosons consist of the vector bosons with spin 1 that mediate the fundamental forces and the scalar Higgs boson with spin 0. All particles of the SM are shown in Fig. 2.1.
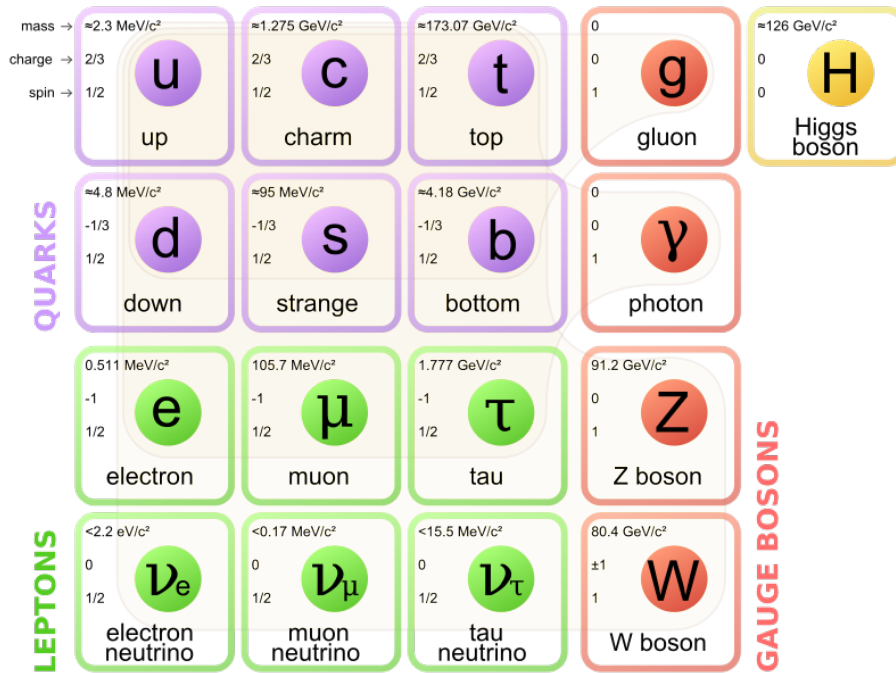
Fig. 2.1: Fundamental particles of the SM [12].

2.1.1 Quarks

Quarks are fermions with spin $1/2$ and exist as three doublets, corresponding to the three generations: up and down (u,d), charm and strange (c,s) and top and bottom (t,b). The mass increases with each generation, where u and d are the lightest quarks. The exact mass is shown in Fig. 2.1. u-type quarks, namely u, t and c, have an electric charge of $+2/3\,e_0$ with e_0 denoting the elementary charge, whereas d-type quarks (d, b, s) have charge $-1/3\,e_0$. The associated anti-quarks have the same properties, but opposite-signed electric charge.

Quarks also have an additional characteristic, the colour charge. The colour charge can occupy three different states: red, green and blue. Antiquarks have the complementary colours. Quarks only form colourless boundary states, having 'white' colour. This can be achieved by three quarks (baryons and antibaryons), consisting of either three quarks

or three antiquarks, one of each colour. Another possibility is a boundary state of two quarks (mesons), consisting of one quark and one antiquark. These two states are well investigated and were considered as the only possible boundary states. Only recently, a new bounded state was discovered at CERN, the pentaquark, which binds five quarks (1 quark and 4 antiquarks, or vice versa) [13]. Bounded states of quarks are called hadrons.

Quarks interact via all three fundamental forces. For the strong and the electromagnetic force the quark flavour (u, d, t, b, c, s) is preserved, whereas for the weak interaction transition between the different quark flavours is allowed.

2.1.2 Leptons

Similar to quarks, leptons are also spin $1/2$ particles and exist in three generations, classified by the electron e, muon μ and tau τ, with the masses shown in Fig. 2.1 and electric charge of $-1\,e_0$. The electron is stable, although for collider experiments the muon can be considered as stable too, as it normally survives until it leaves the detector. On the other hand, the τ has a very short lifetime of approximately $2.9 \cdot 10^{-13}\,\mathrm{s}$, with the consequence that only its decay products can be detected in collider experiments.

For each of these leptons an associated neutrino (ν_e, ν_μ, ν_τ) exists, which forms doublets with the corresponding leptons, e.g. (e, ν_e). The neutrinos have no electrical charge and very low masses, which are almost equal to zero and could not be determined yet. In the SM the lepton flavour is conserved separately for each group, only neutrino oscillations show transition between neutrinos of different generations. Neutrinos only interact via the weak force and are therefore very hard to detect.

For each lepton there also exists an antilepton having opposite charge and chirality.

2.1.3 Fundamental forces and gauge bosons

Tab. 2.1: Properties of the fundamental forces and their gauge bosons.

Fundamental force	Gauge boson	symbol	coupling	mass	charge	spin
Strong force	8 gluons	g	1	0	0	1
Electromagnetism	photon	γ	$\frac{1}{137}$	0	0	1
Weak force	Z boson	Z	10^{-6}	91.2 GeV	0	1
	$W^\pm$ boson	$W^\pm$		80.4 GeV	± 1	1

In the SM the fundamental interactions between particles are described by a gauge theory which is based on local symmetries. When this theory is quantized, the gauge bosons emerge as force carrier between particles. The electroweak force is introduced by a SU(2) x U(1) symmetry denoting the invariance of a Lagrangian under local SU(2), a special unitary group of degree 2, transformation and local U(1), a unitary group of degree 1, transformation. The strong force satisfies a SU(3) symmetry, a local invariance

under SU(3), a special unitary group of degree n, transformation. The different forces interact with different strengths, which are quantified by the coupling constants. These coupling constants are shown in Tab. 2.1 compared to the strength of the strong force, which is set to 1.

The strong force is described by the quantum chromodynamics (QCD). It is mediated by 8 massless gluons, which interact with particles that have colour charge. Therefore gluons only interact with quarks and other gluons, as they also have colour charge. Only considering the gauge theory, fundamental forces are supposed to have infinite interaction ranges and to have massless force carriers. Nevertheless, experiments on the strong force have a finite range. This can be explained by the so-called quark confinement.

The electromagnetic interaction, described by the quantum electrodynamics (QED), is mediated by photons. Photons are massless, like gluons, and interact with all particles that have electric charge: all fermions except the neutrinos and W bosons. In contrast to the strong force, the electromagnetic interaction has infinite range, consistent with a gauge theory.

The force carriers of the weak interaction are three vector bosons with spin 1, namely the Z^0 boson with zero electric charge and the $W^\pm$ bosons with a charge of $\pm e_0$. As mentioned before, a gauge theory implies zero mass for all gauge bosons. Nevertheless, the mediating bosons of the weak force has finite masses. This is introduced to the theory by a local symmetry breaking of the SU(2) x U(1) symmetry. This symmetry breaking requires the coupling to a new scalar field, the Higgs field, that gives mass to the gauge bosons. Due to these masses the weak force has a very short interaction range.

The fourth fundamental force is given by the gravitation, which, together with the electromagnetic force, is the defining force in everyday life as it has infinite range. Nevertheless, on subatomic scales it can be neglected as its strength is much weaker compared to the other forces. It is also the only fundamental force that is not described by the SM. If gravitation can be described using quantum field theory the mediating particle is given by the graviton. So far, the graviton has not been discovered.

2.2 Gauge theory and Higgs mechanism

The SM is described by a relativistic quantum gauge field theory. Historically, the development of the SM began when physicists tried to combine quantum theory with special relativity. The combination of these theories lead to relativistic quantum field theory and its Dirac equation

$$[i\gamma^\mu \partial_\mu - \kappa]\Psi(x) = 0, \qquad (2.1)$$

which describes the free motion of relativistic particles. γ^μ denotes the Dirac matrices, $\Psi(x)$ the wave function of a particle, x the four-vector and $\kappa = \frac{mc^2}{\hbar c}$, in natural units $\kappa = m$, corresponds with $\frac{\kappa}{2\pi}$ to the inverse Compton wavelength with m the particle mass. This theory then was applied to classic electromagnetism, which was only possible

by the introduction of a special local symmetry, gauge invariance. This was the birth of the relativistic quantum gauge field. Local gauge invariance implies the invariance of the Dirac equation under a local transformation of the wave function Ψ

$$\Psi'(x) = e^{iq\Lambda(x)}\Psi(x) = \hat{U}(q)\Psi(x)\,, \tag{2.2}$$

with a parameter q, $\Lambda(x)$ being an arbitrary continuous and differentiable scalar function and $U(q) = e^{iq\Lambda(x)}$ denoting a unitary transformation. Inserting eq. (2.2) into eq. (2.1) leads to a Dirac equation

$$\{\gamma^\mu[i\partial_\mu - q(\partial_\mu\Lambda)] - \kappa\}\Psi(x) = 0\,. \tag{2.3}$$

To obtain the invariance of the Dirac equation, a gauge field is introduced that transforms in the way

$$B_\mu \longrightarrow B'_\mu = B_\mu - \partial_\mu\Lambda\,, \tag{2.4}$$

to compensate the additional term of eq. (2.3). This leads to the equation

$$\{\gamma^\mu[i\partial_\mu - qB_\mu] - \kappa\}\Psi(x) = 0\,. \tag{2.5}$$

Substituting ∂_μ with the covariant derivative

$$D_\mu = \partial_\mu + iqA_\mu\,, \tag{2.6}$$

the Dirac equation has the simple form

$$[i\gamma^\mu D_\mu - \kappa]\Psi(x) = 0\,. \tag{2.7}$$

As stated in the previous chapter gauge theory only leads to massless gauge bosons, which works for photons as stated above but does not explain the masses of the Z and $W^\pm$ bosons of the weak force. This problem is overcome by the Higgs mechanism that introduces the concept of the spontaneous breaking of a local gauge symmetry.

2.2.1 Global symmetry breaking and Goldstone bosons

The dynamics and kinematics of a system are often formulated by a Lagrangian density, which leads to the equations of motion by applying the Euler-Lagrange equations [1]. In the case of global symmetry breaking a complex scalar field

$$\Phi = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2) \tag{2.8}$$

is defined, which has global U(1) symmetry. A possible Lagrangian density of this complex scalar field is

$$\mathcal{L} = (\partial_\mu\Phi^\dagger)\partial^\mu\Phi - m^2\Phi^\dagger\Phi\,. \tag{2.9}$$
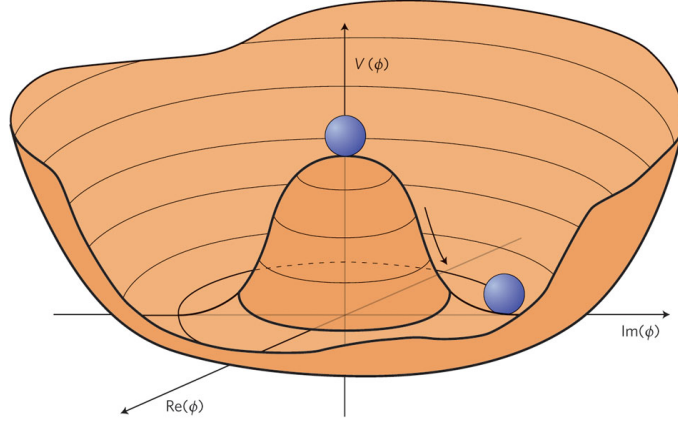
Fig. 2.2: Mexican Hat Potential $V(\Phi^\dagger\Phi)$ with $\Phi = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2)$ being the complex scalar field. The infinite number of ground states are determined by $|\Phi| = \phi_0$ [14].

Considering the field Φ constant in time and space, the derivatives vanish and the minimum energy is given for $\phi_1 = \phi_2 = 0$, which leads to the vacuum state $\Phi = 0$.

In order to obtain a non-vanishing ground state, the Lagrangian density is modified with the potential term $V(\Phi^\dagger\Phi)$

$$V(\Phi^\dagger\Phi) = \frac{m^2}{2\phi_0^2}[\Phi^\dagger\Phi - \phi_0^2]^2\,, \tag{2.10}$$

with ϕ_0^2 being a real parameter. The potential $V(\Phi^\dagger\Phi)$ is illustrated in Fig. 2.2. With the new Lagrangian density

$$\mathcal{L} = (\partial_\mu\Phi^\dagger)\partial^\mu\Phi - V(\Phi^\dagger\Phi)\,, \tag{2.11}$$

the vacuum state is given by $\Phi^\dagger\Phi = \phi_0^2$, which leads to an infinite number of possible states that fulfil the relation $\phi_0 = \sqrt{\phi_1^2 + \phi_2^2}$. However, as soon as the system reaches minimum energy, it occupies one particular state, just as a ferromagnetic system that has one particular direction of magnetization. This means that the ground state does not have the same symmetry any more, the symmetry is spontaneously broken.

If a real vacuum state with $\phi_1 = \phi_0$ and $\phi_2 = 0$ is considered, the U(1) symmetry of the ground state is broken, but the Lagrangian density in eq. (2.11)

$$\Phi \longrightarrow \Phi' = e^{-i\alpha}\Phi\,, \tag{2.12}$$

still has the symmetry. $(\phi_0, 0)$ is now expanded around the ground state, which leads to

$$\Phi = \phi_0 + \sqrt{\frac{1}{2}}(\chi + i\psi)\,. \tag{2.13}$$

Inserting Φ into the Lagrangian density (2.11) gives

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \chi)(\partial^\mu \chi) + \frac{1}{2}(\partial_\mu \psi)(\partial^\mu \psi) - \frac{m^2}{2\phi_0^2}\left[\sqrt{2}\phi_0 \chi + \frac{1}{2}\chi^2 + \frac{1}{2}\right]^2 . \qquad (2.14)$$

The Lagrangian density shows two coupled real fields and can be separated into a free and an interactive term $\mathcal{L} = \mathcal{L}_{free} + \mathcal{L}_{int}$. The free Lagrangian density contains the terms that show only one of the two fields, meaning the terms without interactions

$$\mathcal{L}_{free} = \frac{1}{2}\left[(\partial_\mu \chi)(\partial^\mu \chi) - 2m^2 \chi^2\right] + \frac{1}{2}(\partial_\mu \psi)(\partial^\mu \psi) . \qquad (2.15)$$

The free Lagrangian density contains two fields, one scalar field χ with a mass of $\sqrt{2}m$ and one scalar field ψ without mass, corresponding to a particle with spin 0 and $m = 0$. These massless fields are the result of global symmetry breaking and are called Goldstone bosons. This theorem was introduced by Yoichiro Nambu [15] and elucidated by Jeffrey Goldstone [16].

2.2.2 Local symmetry breaking and the Higgs boson

Now a Lagrangian density is constructed that is invariant under a local U(1) gauge transformation

$$\Phi(x) \longrightarrow \Phi'(x) = e^{-iq\theta(x)}\Phi(x) , \qquad (2.16)$$

where $\theta(x)$ may be time and space dependent. The local U(1) symmetry requires the introduction of a massless gauge field A_μ, which gives a possible Lagrangian density

$$\mathcal{L} = (D_\mu \Phi^\dagger)(D^\mu \Phi) - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - V(\Phi^\dagger \Phi) \qquad (2.17)$$

with

$$D_\mu = \partial_\mu + iqA_\mu \qquad (2.18)$$

denoting the covariant derivation, $F_{\mu\nu}$ being the electromagnetic tensor and $V(\Phi^\dagger \Phi)$ given by the same potential as in eq. (2.10). The Lagrangian density is invariant under the local gauge symmetry

$$\Phi(x) \longrightarrow \Phi'(x) = e^{-iq\theta(x)}\Phi(x) , \qquad (2.19)$$

$$A_\mu(x) \longrightarrow A'_\mu(x) = A_\mu(x) + \partial_\mu \theta(x) \qquad (2.20)$$

The minimum energy of the system is given for a vanishing field A_μ and constant Φ with $\Phi^\dagger \Phi = \phi_0$. Similar to the scalar field in the case of global symmetry breaking, the ground state is also not unique but has an infinite number of possible states. Due to the local symmetry, $\Phi(x)$ can always be chosen real. By setting $\theta(x)$ accordingly, a specific gauge symmetry is is chosen.

Like in the previous chapter, $\Phi(x)$ is again expanded around the real ground state ϕ_0 using the real field $h(x)$

$$\Phi'(x) = \phi_0 + \frac{1}{\sqrt{2}} h(x) \,. \tag{2.21}$$

After inserting $\Phi'(x)$ into the Lagrangian density (2.17) and separating the obtained equation into a free and an interactive term, the free Lagrangian density is given by

$$\mathcal{L}_{free} = \frac{1}{2} \left[(\partial_\mu h)(\partial^\mu h) - 2m^2 h^2 \right] - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} + q^2 \phi_0^2 A_\mu A^\mu \,. \tag{2.22}$$

The free Lagrangian density describes a scalar boson field $h(x)$ with mass $m = \sqrt{2}m$ and a vector field A_μ corresponding to a vector boson with mass $m = \sqrt{2}q\phi_0$. Therefore, the local symmetry gives mass to the gauge bosons, in this toy example the photon A_μ, and leads to a new scalar field $h(x)$ with non-vanishing mass, the Higgs boson. This mechanism was introduced by Peter Higgs [3]. The interactive Lagrangian density contains terms, where the gauge field A^μ mixes with $h(x)$ field.

The electroweak interaction is described by a SU(2) x U(1) symmetry and the spontaneous symmetry breaking of SU(2) x U(1)$_Y$ $\longrightarrow$ U(1)$_{em}$. The derived Lagrangian density can be split into two parts, the

$$\begin{aligned}
\mathcal{L}_1 = & \frac{1}{2} \left[(\partial_\mu h)(\partial^\mu h) - 2m^2 h^2 \right] \\
& - \frac{1}{4} Z_{\mu\nu} Z^{\mu\nu} + \frac{1}{4} \phi_0^2 (g_1^2 + g_2^2) Z_\mu Z^\mu \\
& - \frac{1}{4} A_{\mu\nu} A^{\mu\nu} \\
& - \frac{1}{2} \left[(D_\mu W_\nu^+)^* - (D_\nu W_\mu^+)^* \right] \left[(D^\mu W^{+\nu}) - (D^\nu W^+_\mu) \right] + \frac{1}{2} g_2^2 \phi_0^2 W_\mu^- W^{+\mu} \,.
\end{aligned} \tag{2.23}$$

The first line gives the free massive and neutral scalar Higgs field $h(x)$, the second line corresponds the free massive and neutral vector field $Z_\mu(x)$, the third to the electromagnetic field $A_\mu(x)$ and the last line corresponds to the massive vector fields $W_\mu^+(x)$ and $W_\mu^-(x)$, which interact with the electromagnetic field. The vector fields $Z_\mu(x)$, $W_\mu^+(x)$ and $W_\mu^-(x)$ can be identified with the massive gauge bosons Z, W^+ and W^- and the scalar Higgs field $h(x)$ with the massive Higgs boson H. The remaining terms of the Lagrangian density contain all the other interaction terms and guarantee the renormalization of the theory.

2.2.3 Production modes of the Higgs boson

The SM predicts the dominant production modes at LHC (Fig. 2.3) of the Higgs boson as follows [11].

- The production mode with highest cross section at the LHC is the **gluon-gluon fusion (ggF)**. The Higgs boson is produced via a loop, dominantly, with a running top quark, originating from two gluons (Fig. 2.3a).
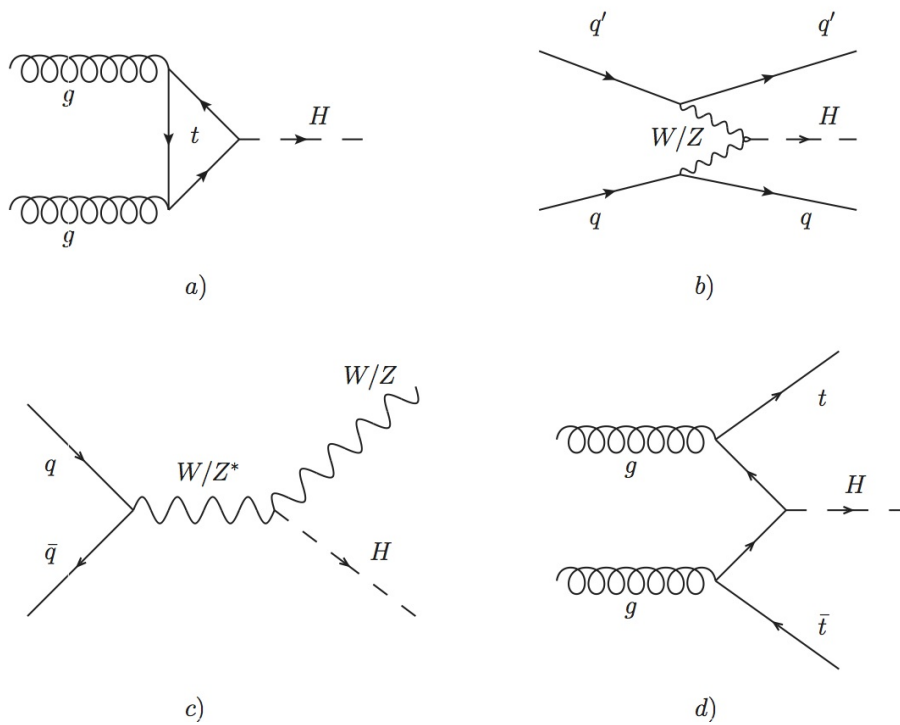
Fig. 2.3: Feynman diagrams for the Higgs production modes, a) gluon-gluon-fusion, b) vector-boson fusion c) Higgs strahlung d) associated production [17].

- The second most probable mode is given by **vector-boson fusion (VBF)** (Fig. 2.3b), which is achieved by two quarks radiating a $W^{\pm}$ or Z^0 boson each which combine to a Higgs boson. This production mode has a very distinct signature, as in addition to the Higgs two hadronic jets can be detected due to the remaining quarks. These jets have a clear angular separation. Due to this unique signature, this production mode is chosen for all analyses carried out through this thesis.

- The third production mode is **Higgs strahlung** (Fig. 2.3c), where the Higgs boson is radiated off a W* or Z* boson. A W or Z boson is then also part of the final state.

- **t$\bar{\text{t}}$ fusion** (Fig. 2.3d), which is like Higgs strahlung an associated production. The Higgs boson is produced together with two top quarks that also appear in the final state.
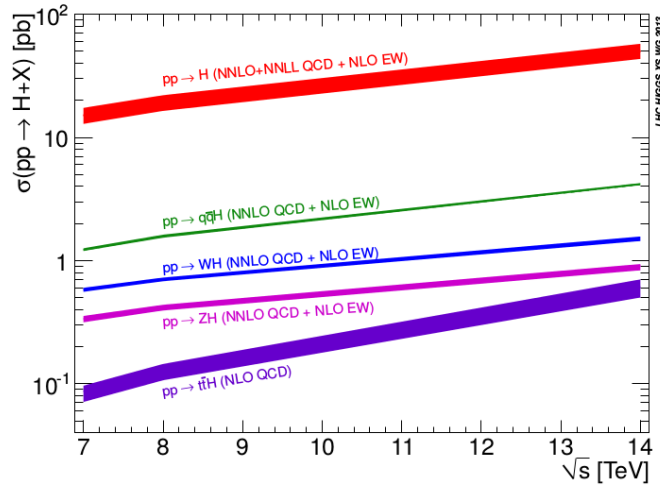
Fig. 2.4: Cross section for the dominant Higgs production modes: Gluon-gluon-fusion ($pp \to H$), Vector-boson fusion ($pp \to q\bar{q}H$), Higgs strahlung ($pp \to WH$, $pp \to ZH$), $t\bar{t}$ fusion ($pp \to t\bar{t}H$) [11].

The cross sections for all these production modes are shown in Fig. 2.4 illustrating the increase of production as a function of the center of mass energy $\sqrt{s}$. Of special interest for the LHC are the cross sections at 7 and 8 TeV for Run I and 13 and 14 TeV for Run II.

2.2.4 Decay modes of the Higgs boson

When a new boson was discovered at CERN by the CMS and ATLAS Collaborations in 2012, its detected properties were found compatible with those of the Higgs boson. As

the mass of the Higgs boson is a free parameter of the SM, after the discovery it was experimentally determined with high accuracy by both collabroations with a combined measured mass of $m_H = 125 \pm 0.21(\text{stat.}) \pm 0.11(\text{syst.})$ GeV [9]. Only when the mass m_H is known, the branching ratios of the predicted decay modes can be calculated, as they strongly depend on the mass of the Higgs boson [18]. This dependency is illustrated in Fig. 2.5. In 2012, the excess in the specific decay modes was most significant in the ZZ and $\gamma\gamma$ modes [7].
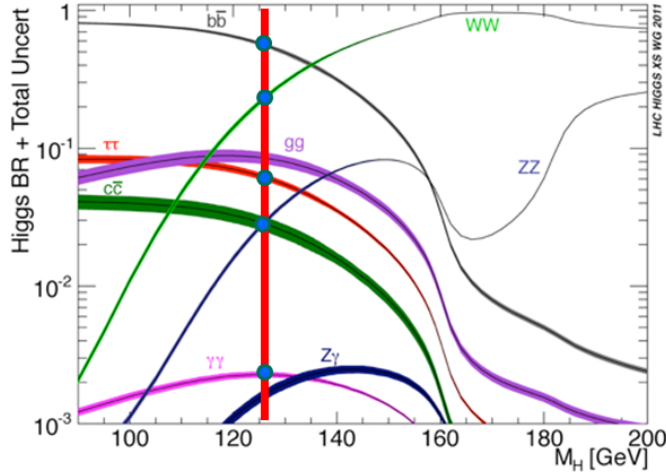
Fig. 2.5: Branching ratio (BR) of the Higgs boson decay modes as a function of the mass of the Higgs boson m_H including the uncertainties [14]

$H \rightarrow \tau^+\tau^-$ decay

The fermions of the SM become massive due to the Yukawa couplings between the Higgs boson and the fermions [10]. Until now, none of the fermionic decay channels are measured with a significance high enough (Z=5) to be declared a discovery, which is necessary for identifying this boson as the SM Higgs boson [11].

The most promising channel is one Higgs boson decaying into two τ leptons with opposite charge ($\tau^+\tau^-$). Compared to other leptonic decays, this channel has a high expected decay ratio in the SM (Fig. 2.5). Only $b\bar{b}$ has a higher branching ratio, but this channel is overwhelmed by the irreducible $p\bar{p} \rightarrow b\bar{b} + X$ background [11].

At the LHC the upper limits for the $H \rightarrow \tau^+\tau^-$ decay channel were placed using pp collision data recorded at $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV. The ATLAS Collaboration found an observed (expected) upper limit of 4.5 (3.4) times $(\sigma * B)_{SM}^{125\,GeV}$ [19], with $(\sigma * B)_{SM}^{125\,GeV}$ denoting the cross section σ times branching ratio B for a SM Higgs boson of $m_H = 125$ GeV in the SM, and the CMS Collaboration placing an upper limit of 3.2 (3.7) times $(\sigma * B)_{SM}^{125\,GeV}$ [20].

In the detector τ leptons are not measured directly, but need to be reconstructed from their decay products. A τ lepton has two leptonic decay channels, namely

$$\tau^- \rightarrow e^- + \bar{\nu}_e + \nu_\tau \,, \tag{2.24}$$

$$\tau^- \rightarrow \mu^- + \bar{\nu}_\mu + \nu_\tau \tag{2.25}$$

and analogously for the antiparticle τ^+. Here, for each decaying τ one light lepton l (e or μ) is detected. Furthermore the τ also decays hadronically in many different decay channels (Tab. 2.2). In this case the τ lepton is reconstructed from 1 or 3 charged particles and several neutral particles, depending on the channel. A reconstructed hadronic decay is denoted as τ_h. The light leptons (e or μ) are symbolized as l, whereas if all leptons are concerned, they are denoted as L (e, μ or τ_h) [21].

Tab. 2.2: Branching ratios of the most relevant τ decay channels [21].

decay channel	BR [%]
leptonic	35.2
$\tau \rightarrow e\nu_e\nu_\tau$	17.8
$\tau \rightarrow e\nu_e\nu_\tau$	17.4
hadronic	64.8
$\tau \rightarrow \pi^\pm \nu_\tau$	11.1
$\tau \rightarrow \pi^0\pi^\pm \nu_\tau$	25.4
$\tau \rightarrow \pi^0\pi^0\pi^\pm \nu_\tau$	9.19
$\tau \rightarrow \pi^0\pi^0\pi^0\pi^\pm \nu_\tau$	1.08
$\tau \rightarrow \pi^\pm\pi^\pm\pi^\pm \nu_\tau$	8.98
$\tau \rightarrow \pi^0\pi^\pm\pi^\pm\pi^\pm \nu_\tau$	4.3
$\tau \rightarrow \pi^0\pi^0\pi^\pm\pi^\pm\pi^\pm \nu_\tau$	0.5
$\tau \rightarrow K^\pm X\nu_\tau$	3.74
others	0.51

Depending on the production mechanism of the Higgs boson (Fig. 2.3), the $H \rightarrow \tau\tau$ decay shows different numbers of charged leptons in the signature. For ggF and VBF only two charged leptons are present, namely LL', denoting the six final states

$$LL' = \mu\tau_h, \ e\tau_h, \ \tau_h\tau_h, \ e\mu, \ \mu\mu, \ ee \,. \tag{2.26}$$

For the associated production with a W/Z boson one or two additional e or μ are present. One additional lepton originates from the W decay, which leads to the four most sensitive final states

$$\begin{aligned} l + L\tau_h =& \mu + \mu\tau_h \,, \\ & e + \mu\tau_h \,/\, \mu + e\tau_h \,, \\ & \mu + \tau_h\tau_h \,, \\ & e + \tau_h\tau_h \,, \end{aligned} \tag{2.27}$$

whereas two additional leptons are present in the Z decay

$$ll + LL' = ll + \mu\tau_h \,,$$
$$+ e\tau_h \,,$$
$$+ e\mu \,,$$
$$+ \tau_h\tau_h \,. \tag{2.28}$$

The ee and $\mu\mu$ channels are not analysed in this content, as these two channels are normally used in the $H \to ZZ \to 4l$ decay channel [20].

In the analysis carried out in the course of this thesis, only the VBF production and the final states $\mu\tau_h$ and $e\tau_h$, with $e\tau_h$ being added secondly, are studied.

Background in the search for $H \to \tau^+\tau^-$

The contribution of different background processes strongly depends on the production mode of the Higgs boson, the decay channel and the final state.

In case of VBF and the final states $\mu\tau_h$ and $e\tau_h$ the biggest irreducible background is given by the Drell-Yan process of a Z/γ^* boson decaying into a pair of τ leptons. Another background leading to the same leptonic final states is $t\bar{t}$ production. Another source of a reducible background is given by a W boson decaying into a lepton l plus neutrino ν and jets, with one jet misidentified as a τ_h. These three backgrounds are investigated in this thesis [20].

3 Experimental setup at CERN

CERN was established in 1954. Nowadays, it is one of the most prestigious research centres for particle physics around the world. It is located near Geneva, Switzerland and currently has 21 member states.

During the last 60 years several groundbreaking achievements have been made at CERN [22], including

- 1960s: Invention of the multiwire proportional chamber by Georges Charpak.

- 1983: Discovery of the W and Z bosons in the UA1 and UA2 experiments using proton-antiproton collisions at the Super Proton Synchrotron (SPS).

- 1989: At the Large Electron-Positron Collider (LEP) the number of light neutrino families was determined to be three.

- 1995: First production of antihydrogen at the Low Energy Antiproton Ring (LEAR).

- 1999: Discovery of direct CP violation in the NA48 experiment at the SPS.

- 2011: Trapping of antihydrogen for 1000 seconds in the ALPHA experiment.

- 2012: Discovery of a boson consistent with the Higgs boson at ATLAS and CMS.

- 2015: Discovery of pentaquarks at LHCb.

3.1 The Large Hadron Collider (LHC)

At the present time the LHC at CERN is the world's largest and most powerful particle collider. It was built after the shut-down of LEP in 2000 by re-using the LEP tunnel with a circumference of 27 km. As the LHC is a particle-particle collider, two rings were set up, in which the beams counter-rotate. For particle-antiparticle colliders, it is sufficient to provide one ring for both beams [23].

One of the main research goals of the LHC is the investigation of electroweak symmetry breaking, which led to the discovery of a Higgs-like boson with a mass of 125 GeV. At the moment, many studies are going on to gain more information on this boson. Further investigations at the LHC include the search for SUSY, extra dimensions, new massive vector bosons, heavy ion physics and SM physics, like top quark physics and B-physics [24].

Tab. 3.1: Accelerator system at CERN [25].

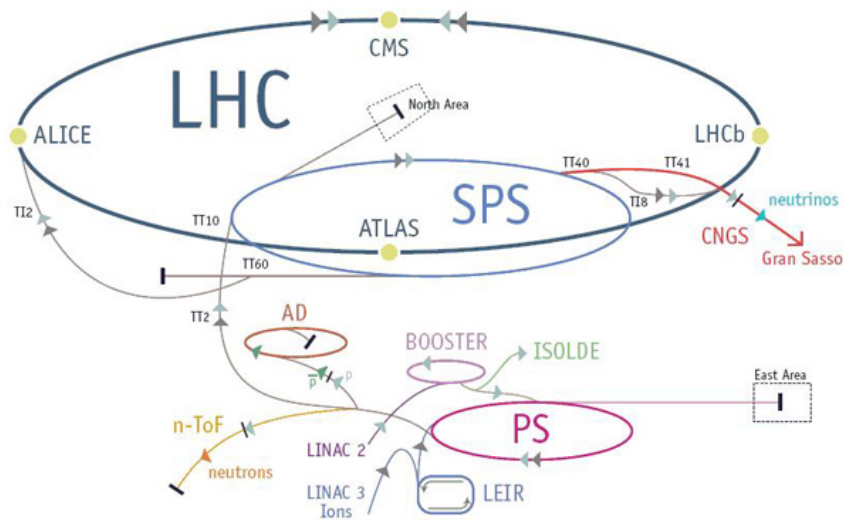| Accelerator	Acronym	max. energy
Linear Accelerator 2	LINAC 2	50 MeV
Proton Synchrotron Booster	PSB	1.4 GeV
Proton Synchrotron	PS	26 GeV
Super Proton Synchrotron	SPS	450 GeV
Large Hadron Collider	LHC	7 TeV

Fig. 3.1: Map of the accelerator complex at CERN [26].

The LHC itself is only the last part of a whole system that combines several particle accelerators. Each accelerator is optimized to its energy range shown in Tab. 3.1. The acceleration in the first four accelerators, before the protons enter the LHC, takes about 20 mins. The final acceleration in the LHC lasts for another 20 mins. Once the acceleration is accomplished, the beams can circulate in the ring for several hours.

In 2010, the first proton collisions at energies of 3.5 TeV per beam were successfully achieved. The combined center-of-mass energy of the two beams was then increased from 7 TeV in 2010 and 2011 to 8 TeV in 2012. After the long shut down from 2013 to 2015, when the LHC was upgraded, the data-taking process started again in 2015 at a center-of-mass energy of 13 TeV.

In addition to the proton-proton collisions, the LHC and its pre-accelerators also provide the possibility of the acceleration of heavy ions, e.g. lead ions. In the final setup, the ions reach energies of 2.76 TeV/u, where u denotes the atomic mass.

The protons do not circulate in the ring in a continuous beam: they are bundled to bunches of around 10^{11} protons. These bunches collide every 25 ns with a luminosity of up to 10^{34} cm^{-2}s^{-1}. The ring of the LHC provides four collision points. At each point one of the four large detectors is situated:

- ALICE, A Large Ion Collider Experiment [27]

- ATLAS, A Toroidal LHC ApparatuS [28]

- CMS, Compact Muon Solenoid [29]

- LHCb, Large Hadron Collider beauty [30]

The LHC setup including all pre-accelerators and detectors is illustrated in Fig. 3.1.

3.2 The CMS detector and its components

The CMS detector, like the ATLAS detector, is one of the LHC multipurpose detectors. It is a combination of many subdetectors that are optimized for the detection of different particles and their properties. Fig. 3.2 gives an overview of the components of CMS. The detector follows an "onion-type" layout and can be divided in different layers.

- The innermost layer includes the **tracking detector** with a high solenoidal magnetic field. It is essential for precise momentum measurements by the curvature of charged particle trajectories, caused by the magnetic field.

- The middle layer contains the **electromagnetic** and the **hadronic calorimeter** and is therefore responsible for the measurement of the energy of the particles. The calorimeters are placed inside the bore of the superconducting solenoid to keep the amount of material in front of the calorimeters as low as possible. The more material particles have to penetrate before they reach the calorimeters, the higher is their energy loss, which leads to uncertainties in the energy measurement.

CMS DETECTOR

Total weight   : 14,000 tonnes
Overall diameter : 15.0 m
Overall length   : 28.7 m
Magnetic field   : 3.8 T

STEEL RETURN YOKE
12,500 tonnes

SILICON TRACKERS
Pixel (100x150 μm) ~16m² ~66M channels
Microstrips (80x180 μm) ~200m² ~9.6M channels

SUPERCONDUCTING SOLENOID
Niobium titanium coil carrying ~18,000A

MUON CHAMBERS
Barrel: 250 Drift Tube, 480 Resistive Plate Chambers
Endcaps: 468 Cathode Strip, 432 Resistive Plate Chambers

PRESHOWER
Silicon strips ~16m² ~137,000 channels

FORWARD CALORIMETER
Steel + Quartz fibres ~2,000 Channels

CRYSTAL
ELECTROMAGNETIC
CALORIMETER (ECAL)
~76,000 scintillating PbWO₄ crystals

HADRON CALORIMETER (HCAL)
Brass + Plastic scintillator ~7,000 channels

Fig. 3.2: Overview of the CMS detector [31].

- The outer layer is given by the **muon detector**. As this layer is shielded by the solenoid and the subdetector systems, only muons reach this detector. Therefore, it is essential for the muon identification and further information of their momenta can be extracted.

In Fig. 3.3 the tracks of different particles in the CMS detector are illustrated. The plot shows the bending of the particles due to the magnetic field and their charge. The plot shows in which subdetector the particles interact and the range of the particles in the CMS detector [24].
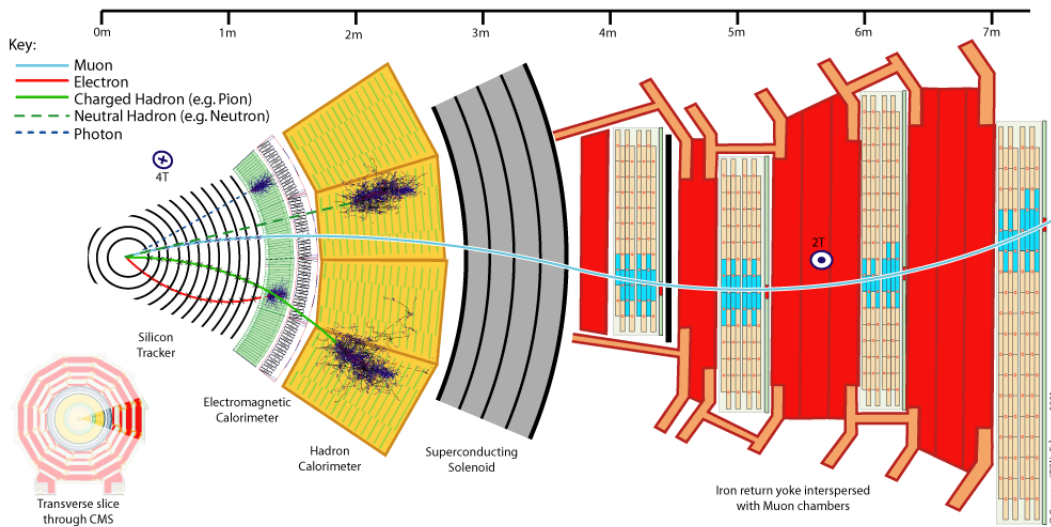
Fig. 3.3: Tracks of different particles in the CMS detector [32].

The CMS coordinate system has its origin centered at the nominal collision point. It is oriented such that the x-axis points inward toward the centre of the LHC ring, the y-axis upward and the z-axis points along in the beam direction to the west. The azimuthal angle ϕ is measured from the x-axis in the xy-plane [29]. The polar angle θ is measured from the z-axis. The pseudorapidity is defined as

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

3.2.1 Superconducting magnet

The choice of a magnet system is a central issue of the detector design. The momentum of a particle is extracted by the Sagitta method. The relative uncertainty is defined by the magnetic field B and the lever arm L,

$$\frac{\sigma_{p_T}}{p_T} \propto \frac{1}{BL^2} . \tag{3.2}$$

The design goal is that relative uncertainty on the transverse momentum p_T of muons of approximately 1 TeV should not exceed 10%. This is achieved by either a high magnetic field, which is the case for CMS, or by a big detector, which was chosen for the ATLAS detector. Using a superconducting solenoid, a magnetic field of 4 T is reached. The solenoid shows a length of 12.9 m and an inner bore of 5.9 m. In this bore, the tracker and the two calorimeters are situated [24, 29].

3.2.2 Inner tracking system

The innermost layer of the CMS detector is the silicon tracker, which needs to satisfy several requirements for the measurement process.

- Considering the luminosity of the LHC, approximately 1000 charged particles per bunch-crossing are expected. Together with the bunch crossings every 25 ns this requires high granularity and fast response of the tracker.

- As particles with a transverse momentum of $p_T > 1$ GeV are of special interest for most searches, the tracker has to have efficient reconstruction of these particles.

- During each bunch crossing, several collisions take place, called pileup. Pileup is a problem for the reconstruction, hence the identification of pileup vertices must have a high resolution.

- Furthermore, a good coverage of all particles up to a pseudorapidity of $\eta \approx 2.5$ is desired, with θ denoting the polar angle of the particle direction. An event can only be reconstructed correctly, if all decay products are detected.

To reach enough granularity and the required precision, the tracker is constructed with 10 layers of silicon microstrip detectors. To enhance the measurement precision on the position of secondary vertices, 3 layers of silicon pixel detectors are located close to the collision region [24, 29].

3.2.3 Electromagnetic and hadronic calorimeter

The aim of these two calorimeters is a precise measurement of the particle energy. The electromagnetic calorimeter (ECAL) is most sensitive to photons and electrons, whereas the hadronic calorimeter (HCAL) reconstructs also part of the energy of hadrons.

The ECAL consists of lead tungstate scintillating crystals. The granularity is given by the size of the crystals. These crystals show fast light emission (80% within 25 ns) and are radiation-hard. In the barrel region ($|\eta| < 1.479$), the light is detected by avalanche photodiodes (APD). In the endcap region, vacuum phototriodes are installed that cover the pseudorapidity range $1.479 < |\eta| < 3$. The ECAL structure is illustrated in Fig. 3.4 [24, 29].

The HCAL surrounds the ECAL. To enable the reconstruction of the missing transverse energy E_T^{miss} a good containment and hermeticity is required. Hence, the amount
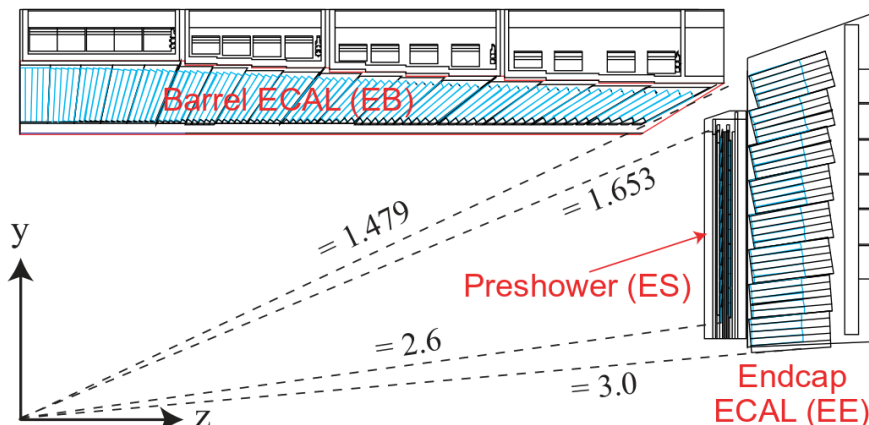
Fig. 3.4: Geometrical configuration of the ECAL, transverse section [24].

of absorber material of the HCAL inside the solenoid is maximised. The absorber material is chosen to be brass complemented by a scintillator layer. The HCAL is divided into four subsystems. The hadronic barrel (HB) and endcap (HE) calorimeters sit behind the ECAL as seen from the interaction point covering a range $|\eta| < 1.3$ and $1.3 < |\eta| < 3$, respectively. The hadronic outer calorimeter (HO) is placed outside the magnetic coil covering the same pseudorapidity range $|\eta| < 1.3$ as the HB. An extended range $3 < |\eta| < 5$ is covered by the hadronic forward calorimeters (HF). The locations of the subsystems in the CMS detector are shown in Fig. 3.5 [24, 29].

3.2.4 Muon detector

Muons produced in the central interaction region are measured twice in the CMS detector, in the inner tracker system and in the muon detector. The muon detector is essential for the identification of the muon, but does not give a satisfying resolution on the transverse momentum p_T by its own. To increase the resolution the measurements of the inner tracker and the muon detector are combined.

The muon detector is divided into three regions: the barrel region ($|\eta| < 1.2$) and the two endcap regions ($0.9 < |\eta| < 2.4$). These regions are subject to different levels of radiation, which influence the choice of the detector technologies. In total, three different gaseous detectors are used. In the barrel region, the neutron-induced background rate as well as the magnetic field and the expected muon rate are low, which led to the choice of a drift tube (DT) chamber. In the endcap regions, where the expected muon rate is high as well as the neutron-induced background and the magnetic field, cathode strip chambers (CSC) are installed. In addition, resistive plate chambers (RPC) are used in all regions. In Fig. 3.6 the layout of the muon system is illustrated [24, 29].
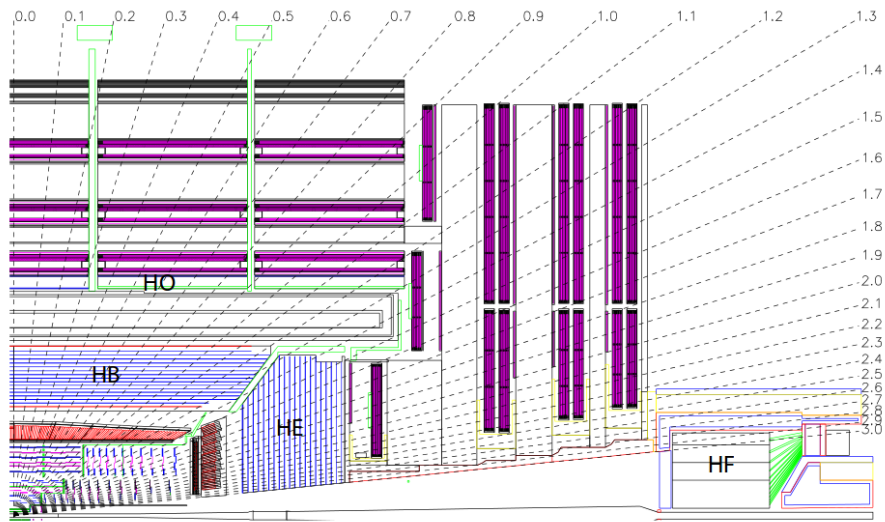
Fig. 3.5: CMS detector showing the locations of the hadronic barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters, longitudinal view [29].
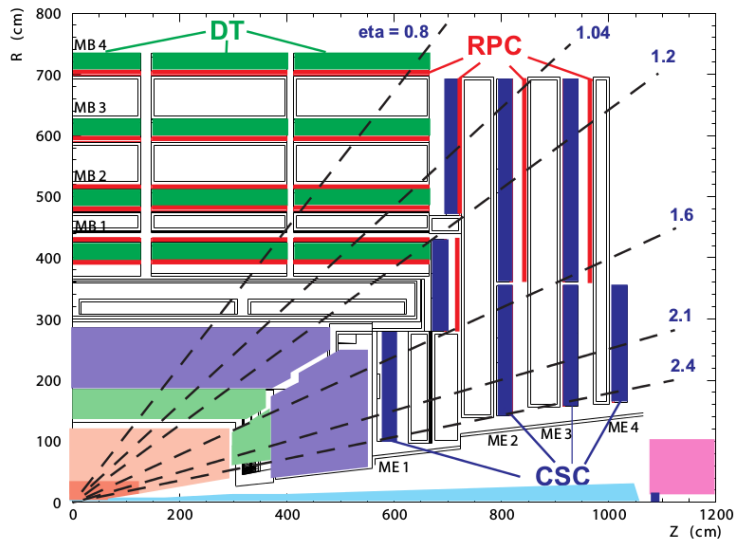
Fig. 3.6: Layout of the muon system for one quarter of the CMS detector [24].

3.2.5 Trigger and data acquisition (DAQ)

Taking into account the design luminosity of the LHC and the bunch crossing rate of 40 MHz, up to 10^9 interactions are expected each second. It is not possible to read out and store this amount of events since 1 event accounts to ≈ 1 MB, which would lead to a data rate of 60 TB/s. A trigger selects the potentially interesting events and reduces the rate of recorded events to a few hundred per second. This corresponds to a reduction factor of 10^6.

The trigger and data acquisition (DAQ) system of CMS combines four different parts: the electronics in the detection, the Level-1 trigger (L1), the readout network and the High-Level trigger (HL T), which corresponds to an online event filter that is performed in the processor farm. More detailed information on the trigger and DAQ can be found in Refs. [24] and [29].

4 Analysis methods and boosted decision trees

The goal of experimental high energy physics is the observation of specific particle interaction processes. This involves the analysis of large data samples. A successful interpretation of this data requires advanced computer techniques. In the case of the LHC, the samples consist of proton-proton collision events. Each event contains the reconstruction of the particles that are produced during the interaction process and detected, e.g. electrons, muons or jets. A jet corresponds to a collimated spray of hadrons or other particles caused by the hadronization of quarks or gluons. Different physical processes have different signatures in the event, like the number of jets or leptons, or specific kinematic properties. The computer algorithm used for the analysis has to perform an event classification and efficiently separate signal events from background events based on the measured event properties. Variables with high separation power are defined as input variables, which are used for the event classification. The choice of these variables strongly depends on the processes that are included in the analysis.

This chapter provides two different analysis approaches, a cut-based analysis and a multivariate analysis based on machine learning techniques. Since the focus of this thesis lies on BDTs, this event classification method is described in more detail.

4.1 Cut-based analysis

The most straight forward way to separate signal and background is a simple cut-based analysis. In this case a separation value, which corresponds to a cut between signal and background region, is set for each input variable. Starting with the cut on the first variable, only signal-like events, meaning events lying on the side of the cut with more signal events, survive and therefore proceed to the next separation cut on the second variable, and so on. The events on the "background-side" of the cut are removed and not considered for evaluation. The aim is to keep as many signal events and remove as much background as possible. The critical point consists in finding the optimal cuts, which can be performed using the TMVA package [33].

During this process, two problems emerge. Firstly, there is a probability for background events to survive this separation process. This leads to background events that are wrongly classified as signal and therefore wind up in a loss of purity of signal events. Secondly, it is possible that signal events do not survive the selection and are wrongly classified as background. This leads to a decrease in efficiency.

One way to reduce the impact of these two problems is to extract more information from the input variables, which can e.g. be accomplished by considering correlations between this variables and the simultaneous analysis of information. Unfortunately, this is not straight-forward for a pure cut-based method. As a consequence multivariate analysis methods (MVA) based on machine learning are used.

4.2 Multivariate analysis

It is often not sufficient to consider variables one after the other. Analysis methods that analyse several variables simultaneously often perform better. Multivariate analyses (MVA) fulfil this requirement.

MVAs provide the possibility of simultaneous observation of several variables, which means that relations between variables have an impact on the outcome. The information extracted from these variables is then combined to one single classification quantity.

4.2.1 Training

In high energy physics, supervised learning is in wide use. In the course of this the analysis method is built using a large number of pre-classified events, meaning the method is 'trained'. These classified events often correspond to samples that are produced by a Monte Carlo simulation. For limited sample size, bigger training samples lead to better outcome. Nevertheless, at some point the training performance reaches saturation that can not be enhanced any further. During the training, the method learns the properties and peculiarities of the samples and how to distinguish between signal and background [33].

4.2.2 Testing

Typically, MVA methods assign a continuous value (MVA score x) of a certain interval, e.g. [-1;1], with $x < 0$ corresponding to a background-like event and $x > 0$ indicating a signal-like event. A well-performing algorithm is characterized by a high separation power between signal and background events, i.e. few signal events in the background.

As the final goal of a MVA is the event classification of data, the MVA is tested on an independent second MC sample and is optimized with respect to the performance for this unknown sample. Ideally, the MVA shows a similar performance as for the training samples with only slight deviations due to statistical fluctuations. In most cases, the MVA performs worse for the second sample compared to the training sample. This is caused by the so-called overtraining.

4.2.3 Overtraining

Overtraining is a problem caused by a training sample that is too small. It occurs if the machine learning problem has many free parameters but the samples do not provide enough independent and general data points. In this case, statistical fluctuations in the training sample and not general properties determine these parameters. With respect to overtraining, it can be of advantage to reduce the number of free parameters in order to achieve a better performance of the method. Overtraining is not a problem per se as long as it is considered during the optimization process and will always occur to some extant. It only gets problematical, when it causes the performance of a MVA method to decrease significantly compared to other methods.

4.3 TMVA - Toolkit for Multivariate Analysis

The TMVA toolkit, which is a ROOT-integrated environment [34], provides a possibility for the processing and application of multivariate classifiers. The toolkit includes several different multivariate methods, which are all using a training sample (already classified events) to derive the mapping function used for classification. TMVA was initially designed for high energy physics [33].

4.4 Boosted decision trees (BDT)

There are various different MVA methods that can be used for signal extraction. This thesis focuses on Boosted Decision Trees.

A decision tree consists of a sequence of yes or no decisions that can be illustrated as a two-dimensional tree structure. Starting from the root node (Fig. 4.1) the variable with best separation power and its value is evaluated during the training process (growing of the decision tree). The separation splits the phase space into two subsets and the process is then repeated for each set. After several turns, the final leaf nodes are reached, which are classified as background- or signal-like. The classification depends on the number of training events ending up in each particular node. If a majority of signal events is observed, the leaf node will be classified as signal, otherwise as background. The main difference between the cut-based method and BDTs is that in case of cut-based, only one small part of the whole phase space is evaluated, while for BDTs all subspaces of the whole phase space are classified as signal- or background-like and included in the evaluation.

The 'Boosting' refers to the extension from one tree to many trees. Each tree is created by the same training sample, but using differently weighted events. If in the first tree a signal event ends up in a background node, it gets assigned a higher weight for the generation of the second tree. This boosting significantly increases the performance of the analysis method. In the end, all derived trees are combined to one single classifier.
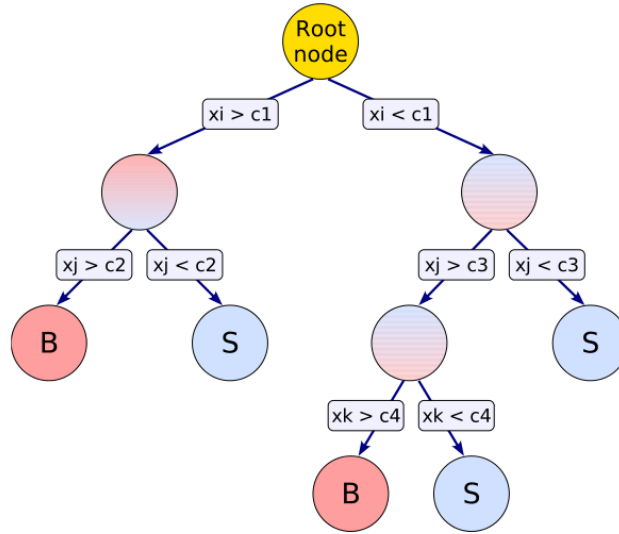
Fig. 4.1: Schematic view of a decision tree [33].

The performance of the combined classifier is given by the weighted average of the performance of the single trees.

Naturally, this method needs to be adapted to the particular problem that needs to be solved. The BDT method implemented in the TMVA package provides several parameters that characterize the BDT that will be constructed during the training process. The parameters with biggest impact on the performance of the analysis are listed:

- **Shrinkage**, corresponding to the learning rate of the training process, which determines how fast the BDT adapts to the training sample and therefore also to fluctuations.

- **NTrees**, which is the number of created trees.

- **MaxDepth**, the maximal number of yes-no questions a tree can contain

- **nCuts**, giving the number of grid points in the variable range which are used for finding the optimal cut

- **BaggedSampleFraction**, denoting the relative size of the random subsample of the training sample used during the training process

- **MinNodeSize**, the minimum percentage of the training events that must end up in each final leave node

- **BoostType**, declaring, which boosting algorithm is used (e.g. AdaBoost, GradBoost)

4.4.1 Boosting algorithms

The most commonly used boosting algorithm is the adaptive boosting (AdaBoost) [33]. For each final node of a decision tree, a misclassification rate ϵ is defined, which is the ratio of the number of misidentified events to the total number of events in each final leaf node. For events that get misclassified in one decision tree a boost weight α is applied, which is multiplied with the older weights, and is used for growing the next tree. The boost weight is set by the misclassification rate ϵ for signal or background, respectively, and is given by

$$\alpha = \frac{1 - \epsilon}{\epsilon} \, . \tag{4.1}$$

After all events got assigned a new weight, the events are renormalized to the initial number of expected events.

The result of an individual classifier (one tree) for each event is defined as $y(\mathbf{x})$, where $\mathbf{x}$ denotes the tuple of input variables. $y(\mathbf{x})$ can only have the values $+1$ or -1, with $y(\mathbf{x}) = +1$ when the event ended up in a signal node or $y(\mathbf{x}) = -1$ for events that ended up in background nodes, respectively. The classification result for each event of the boosted classifier is then given by

$$y_{Boost}(\mathbf{x}) = \frac{1}{N_{classifiers}} \cdot \sum_{i=1}^{N_{classifiers}} \left[\ln(\alpha_i) + y_i(\mathbf{x}) \right] \, , \tag{4.2}$$

with $N_{classifiers}$ corresponding to the number of individual classifiers (number of trees), α_i to the misclassification rate assigned to the event in tree i and $y_i(\mathbf{x})$ to the classification result of the particular tree encoded as ± 1. A high (low) $y_{Boost}(\mathbf{x})$ corresponds to a signal-like (background-like) event.

The adaptive boost performs best for weak individual classifiers, so-called 'weak learners'. For decision trees, this requirement corresponds to small trees, meaning a small depth of 2 to 3. In the analyses carried out throughout this thesis, this is fulfilled for all BDTs.

The gradient boosting (GradBoost) [33], which will be used in the analysis later on, can be understood as an expansion of the adaptive boost. Given a training sample of N events, each event is characterized by a known variable set $\{y_i, \mathbf{x}_i\}$ with $i = 1, ..., N$. y_i corresponds to the encoded process of event i, either $+1$ for signal or -1 for background, and $\mathbf{x}_i$ to the set of input variables of the i-th event. The goal of the gradient boosting is to find a function $F^*(\mathbf{x})$ that maps y to $\mathbf{x}$, such that it minimizes some specified loss function $L(F(\mathbf{x}), y)$.

Starting from an initial guess $F_0(\mathbf{x})$ the recursive rule

$$F_m(\mathbf{x}) = F_{m-1}(\mathbf{x}) + \beta_m h(\mathbf{x}; \mathbf{a}_m) \, , \tag{4.3}$$

with the parameter set $(\beta_m; \mathbf{a}_m)$, is applied M times. $h(\mathbf{x}; \mathbf{a}_m)$, characteristic for the m-th weak learner, denotes a simple function of the input variables $\mathbf{x}$ with the parameters

$\mathbf{a}_m = \{a_1, .., a_n\}_m$. For each iteration m, the parameters $\mathbf{a}_m$ and the weighting factor β_m are chosen such that the sum of the loss functions over all events is minimized

$$(\beta_m, \mathbf{a}) = \underset{\beta, \mathbf{a}}{\text{argmin}} \sum_{i=1}^{N} L(F_{m-1}(\mathbf{x}_i) + \beta h(\mathbf{x}_i; \mathbf{a}), y_i). \tag{4.4}$$

For the adaptive boost an exponential loss function $L(F, y) = e^{-F(\mathbf{x})}$ is used, which leads to the algorithm stated above. The disadvantage of this loss function lies in the lack of robustness in the presence of outliers. Thus in noisy settings, the performance of the adaptive boost is expected to decrease.

In the TMVA version used, the loss function of the gradient boost is given by the binomial log-likelihood loss

$$L(F, y) = \ln(1 + e^{-2F(\mathbf{x})y}). \tag{4.5}$$

As this loss function can not be minimized straightforwardly, a steepest-descent approach is chosen. Thus the current gradient of the loss function is calculated, followed by growing a regression tree whose leaf values are adjusted to the mean of the gradient in each region. Iterating this procedure minimizes the loss-function as required [35].

Like the adaptive boost, the gradient boost performs best for weak classifiers. As additional benefit, these small trees are also less prone to overtraining compared to single decision trees without boosting [33] .

4.4.2 Performance of BDTs

Compared to other MVA methods, BDTs are often referred to as the best 'out of the box' classifier as only little tuning is needed to obtain a high performance. This is due to the simplicity of the method: only one-dimensional optimization is required. This also leads to relatively low computation time for the training compared to other MVAs.

A single decision tree strongly depends on fluctuations and can lead to significant variations in performance. Using a boosted decision tree with a high number of trees, but smaller trees and a low learning rate, this effect can be reduced significantly.

Not only the parameters of the BDT affect the performance, but also the number of input variables. In contrast to many other MVA methods, input variables that do not have a strong separation power do not have a strong influence on the outcome of the BDT analysis. This is due to the way a BDT is constructed, as only the strongest variables are chosen for the cuts.

Nevertheless, the theoretically best performance that can be obtained by BDTs is significantly below the theoretical possibilities of other MVAs. This is mostly due to the simplicity of the method. Despite this, in most cases BDTs still show better performance than other MVAs. Most other MVAs require a very big number of training events, which is often not available.

5 Quantification of the performance

In order to choose the most powerful classification algorithm for a present problem, a method to quantify the performance of different algorithms is needed. For comparing different boosted decision trees the loss function (see previous chapter) provides a possibility. Nevertheless, as soon as different analysis methods need to be compared, no matter whether multivariate or not, a different quantity must be set as Figure of Merit (FoM). Among the various options, two will be discussed in this chapter, namely the approximate median significance (AMS) and the p-value with its corresponding Gaussian significance Z.

Before the quantification of the performance of the event classification can take place, the information that is provided by different analysis methods needs to be preprocessed.

5.1 Preparation of the analysis output before quantification of the performance

During the classification procedure performed by a BDT, which corresponds to the main analysis method throughout this thesis, each event gets assigned a classification value (BDT score) between 1 (signal-like) and -1 (background-like). Two histograms, one for signal and one for background, with the same binning are filled with the obtained BDT scores.

In the beginning the binning is chosen equidistantly and rather narrow to preserve as much information as possible. In order to reduce the impact of statistical fluctuations on the FoM, a rebinning is performed with respect to the statistical uncertainties and the amount of background events in the bins. A value for the maximal statistical uncertainty per bin (RELSTATMAX) is defined that must not be exceeded, as well as a background regulation term (BINC) that determines the factor of the increase in background events per bin from signal region to background region. Starting on the right side of the histogram (signal region) the rebinning algorithm calculates the statistical uncertainty of the furthest to the right bin. If this value exceeds the limit set by RELSTATMAX, the second furthest bin is added. The calculation is performed again and the procedure is repeated until the statistical requirement is fulfilled. Then the process is performed for the second bin, but from here on also a second requirement needs to be satisfied, which is defined by the BINC parameter. This parameter corresponds to the factor of increase in background events per bin.

These two parameters can be interpreted to following way. A decrease of RELSTAT-

MAX leads to lower statistical uncertainties and a higher number of signal events in the first bins in the signal region, but also to more background events. In case of BINC, an increase of the value leads to more background events per bin especially in the background region and therefore also results in a reduction of the total number of bins. After this 'smart rebinning' is performed, the extracted histogram is ready for quantification.

In contrast to the BDT method, where each event that is evaluated gets assigned a continuous value, no comparable quantity exists for the cut-based method. Instead, only events that satisfy several requirements survive the analysis and are available for quantification. To estimate the performance of these classification methods the output of the analysis can be processed in order to deduce a histogram showing a distribution of the survived signal and background events. The higher the separation between signal and background events in this distribution, the better. A commonly used distribution variable for the cut-based analysis is given by $m_{\tau\tau}^{SVFit}$, denoting the SVFit mass estimation of the $\tau\tau$-system. The histogram is then used for the quantification of the separation power of the cut-based analysis.

5.2 Approximate median significance (AMS)

For quantification of the performance of the event classification and signal extraction a histogram showing a distribution of signal and background events is used. The histogram can be derived as discussed in sec. 5.1.

For each bin i of this histogram the approximate median significance AMS_i is calculated by

$$AMS_i(s_i, b_i) = \sqrt{2\left((s_i + b_i + b_{reg}) \cdot \ln(1 + \frac{s_i}{b_i + b_{reg}}) - s_i\right)} \qquad (5.1)$$

with s_i being the number of signal events and b_i the number of background events in bin i [36]. The Figure of Merit, in this case the total AMS, is then given by the square root of the quadratic sum of the approximate median significance of each bin of the histogram, namely

$$AMS = \sqrt{\sum_{bins} AMS_i^2(s_i, b_i)}. \qquad (5.2)$$

The parameter b_{reg} shown in eq. (5.1) corresponds to a regulation term, which disfavors results with low event counts. For tuning the BDT parameters and picking the right working point during the BDT optimization, b_{reg} is chosen as 1. This prevents choosing a non stable working point that only reaches a higher AMS due to a very low background in single bins caused by fluctuations. For quantifying the performance to compare BDTs for different analyses, the regulation term is set to a very low value, as a higher b_{reg} would underestimate the real signal extraction efficiency. For $s << b$ eq.~(5.1) reduces to $\frac{s}{\sqrt{b}}$.

In the actual calculation of the AMS also statistical and systematic uncertainties are included, which are left out in eq. (5.1) for simplicity.

5.3 Significance and p-value

The p-value is extracted by testing a statistical hypothesis against an observed measurement. It gives the probability of obtaining another measurement with equal or even higher incompatibility while the tested hypothesis is true. Evaluating the p-value can be based on the number of signal events found in a specific region or the corresponding signal-to-background likelihood ratio [37].

Once the p-value is known, it can be converted into a significance Z in the following way (see Fig. 5.1 for illustration).
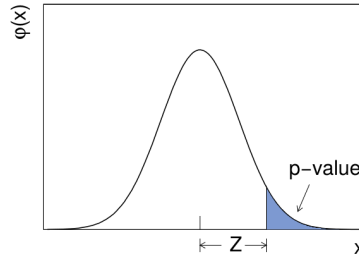
Fig. 5.1: Standard Gaussian distribution, $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, showing the relation between the p-value and the significance Z [37].

Considering a standard Gaussian distribution, the Gaussian significance Z is given by

$$Z = \Phi^{-1}(1 - p) \,, \tag{5.3}$$

with the integral of the upper tail of the standard Gaussian distribution $\varphi(x)$ corresponding to the probability p, namely

$$p = \int_Z^\infty \varphi(x) \, dx \,, \tag{5.4}$$

and Φ^{-1} denoting the inverse function of

$$\Phi(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^Z e^{-\frac{x^2}{2}} \, dx \,. \tag{5.5}$$

In order to declare an observed excess a discovery, a frequentist statistical test is carried out. First of all, a null hypothesis H_0 is declared, which only includes the background events. Then one also needs to specify a second hypothesis H_1, including both background and signal (MC sample or data), which is tested against H_0. Depending

on the agreement of these two hypotheses, the p-value is evaluated, which is equal to the probability of finding data with equal or greater incompatibility. If the p-value is small enough, the hypothesis H_0 can be excluded and a discovery is made. In common use the background-only hypothesis is rejected for a Gaussian significance of at least $Z = 5$, which is equal to a p-value of $2.87 * 10^{-7}$. For the rejection of a signal hypothesis the threshold is normally set to p=0.05, meaning 95% confidence level and a significance of Z=1.68.

In a frequentist statistical test additional to some parameters of interest, e.g. cross section, nuisance parameters can be included. The values of these nuisance parameters is not known a priori, but can be fitted from data. This additional freedom introduced to the model provides a sufficient flexibility to adapt to the impact of systematic uncertainties, but hence also leads to a decrease in sensitivity.

In order to illustrate the process of determining the p-value one example is discussed in more detail. In an experiment, for each event one or more kinematic variables are measured and displayed in histograms. For a variable x the entries of all bins of the histogram are given by

$$n = (n_1, ..., n_N),$$ (5.6)

with N denoting the number of bins. The expected number of events in each bin corresponds to

$$E[n_i] = \mu s_i + b_i,$$ (5.7)

with μ corresponding to the signal strength and with s_i and b_i being the expected numbers for signal and background events in the i-th bin, namely

$$s_i = s_{tot} \int_{bin\ i} f_s(x; \theta_s)\, dx,$$ (5.8)

$$b_i = b_{tot} \int_{bin\ i} f_b(x; \theta_b)\, dx.$$ (5.9)

where $f_s(x; \theta_s)$ and $f_b(x; \theta_b)$ are the probability density functions for signal and background. If $\mu = 0$, one gets the background-only hypothesis H_0. Moreover, s_i and b_i depend on their nuisance parameters θ_s and θ_b, which from now on will be referred to as $\theta = \theta(\theta_s, \theta_b)$.

The likelihood function that is needed for evaluating the p-value is given by the product of the Poisson probabilities for all bins

$$L(\mu; \theta) = \prod_{i=1}^{N} \frac{(\mu s_i + b_i)^{n_i}}{n_i!} e^{-(\mu s_i + b_i)} \prod_{k=1}^{M} \frac{u_k^{m_k}}{m_k!} e^{-u_k}.$$ (5.10)

The second product corresponds to the impact of the nuisance parameter. m_k and u_k are derived by the use of a control sample, where mainly background events are expected. m_k corresponds to the number of events in each of the M bins and $u_k = u_k(\theta)$ is the expectation value of m_k, a calculable quantity that depends on the nuisance parameters.

The profile likelihood ratio for testing a hypothesized value μ is then given by

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\theta}}(\mu))}{L(\hat{\mu}, \hat{\theta})}, \tag{5.11}$$

where $L(\hat{\mu}, \hat{\theta})$ is the (unconditionally) maximized likelihood function, with $\hat{\mu}$ and $\hat{\theta}$ being the maximum-likelihood (ML) estimators, and $L(\mu, \hat{\hat{\theta}})$ is the maximized function for given μ (conditional), with $\hat{\hat{\theta}}(\mu)$ being the conditional ML estimator of θ for given μ.

As one can see in equation (5.11) λ fulfils $0 \leq \lambda \leq 1$. The higher λ, the better the agreement between data and the hypothesized value of μ. The nuisance parameters broaden the profile likelihood ratio, which reflects the impact of systematic uncertainties.

Instead of λ,

$$t_\mu = -2 \ln \lambda(\mu) \tag{5.12}$$

is commonly used for a statistical test. The higher t_μ, the stronger the disagreement between the data and the hypothesized μ. The quantification of the incompatibility is given by the p-value

$$p_\mu = \int_{t_{\mu,obs}}^{\infty} f(t_\mu|\mu) \, dt_\mu \,, \tag{5.13}$$

where $t_{\mu,obs}$ denotes the observed value of t_μ from data and $f(t_\mu|\mu)$ corresponds to the probability density function (pdf) under assumption of a signal strength μ (see Fig. 5.2 for illustration). The Gaussian significance can then be derived analogously to eq. (5.3).
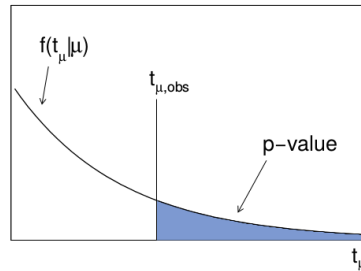
Fig. 5.2: Probability density function (pdf), showing the relation between the observed value $t_{\mu,obs}$ from the test statistics and the p-value [37].

In case μ is assumed non-negative, the definition of λ needs to be adopted to

$$\tilde{\lambda}(\mu) = \begin{cases} \frac{L(\mu, \hat{\hat{\theta}}(\mu))}{L(\hat{\mu}, \hat{\theta})} & \hat{\mu} > 0 \\ \frac{L(\mu, \hat{\hat{\theta}}(\mu))}{L(0, \hat{\hat{\theta}}(0))} & \hat{\mu} \leq 0 \end{cases}, \tag{5.14}$$

in order to avoid problems if the data leads to $\hat{\mu} \leq 0$. $\hat{\hat{\theta}}(\mu)$ and $\hat{\hat{\theta}}(0)$ refer to the conditional ML estimators with a signal strength of μ. Thus $\hat{\lambda}(\mu)$ is used instead of $\lambda(\mu)$

in eq. (5.12) leading to

$$\tilde{t}_\mu = -2\ln\tilde{\lambda}(\mu)\,. \tag{5.15}$$

In case of trying to make a discovery of a signal process, the aim is to reject the background-only hypothesis with a certain p-value. Considering eq. (5.7) the background-only hypothesis corresponds to $\mu = 0$. As this is the most important case in this analysis the notation $q_0 = \tilde{t}_0$ is introduced. With $\mu = 0$ eq. (5.14) reduces to

$$\tilde{\lambda}(0) = \begin{cases} \frac{L(\mu,\hat{\hat{\theta}}(\mu))}{L(0,\hat{\theta}(0))} & \hat{\mu} > 0 \\ 1 & \hat{\mu} \le 0 \end{cases}, \tag{5.16}$$

and eq. (5.15) then leads to

$$q_0 = \begin{cases} -2\ln\lambda(0) & \hat{\mu} > 0 \\ 0 & \hat{\mu} \le 0 \end{cases}. \tag{5.17}$$

It is important to keep in mind that it is assumed that the existence of the signal process only leads to an increase in events ($\hat{\mu} > 0$), never a decrease. This also means that a $\hat{\mu}$ value below 0 does show incompatibility with the background-only hypothesis, not due to signal events, but rather gives evidence for a fluctuation or underestimated systematic errors in the background. In this analysis it is assumed that all systematic uncertainties are already considered by the nuisance parameters θ.

A low test statistic q_0 for $\mu = 0$ shows a very good agreement with the background-only hypotheses, meaning that it is extremely unlikely that a signal is present. The higher q_0 gets, the smaller the agreement with the hypothesis and the more significant the signal. To quantify the disagreement, the p-value is computed by

$$p_0 = \int_{q_{0,obs}}^{\infty} f(q_0|0)dq_0\,, \tag{5.18}$$

with $f(q_0|0)$ denoting the pdf of the statistic q_o under assumption of the background-only hypothesis $\mu = 0$.

Now the probability density function (pdf) for the case of discovery will be discussed. The approximation

$$-2\ln\lambda(\mu) = \frac{(\mu - \hat{\mu})^2}{\sigma^2} + \mathcal{O}\left(\frac{1}{\sqrt{N}}\right)\,, \tag{5.19}$$

which is valid in the large sample limit and for one parameter of interest, is applied, leading to

$$q_0 = \begin{cases} \frac{\hat{\mu}^2}{\sigma^2} & \hat{\mu} > 0 \\ 0 & \hat{\mu} \le 0 \end{cases}, \tag{5.20}$$

where $\hat{\mu}$ follows the Gaussian distribution with a mean $\mu' = 0$ and standard deviation σ. σ is derived from the covariance matrix of the estimators of all nuisance parameters θ_i,

$$V_{ij} = \mathrm{cov}[\hat{\theta}_i, \hat{\theta}_j] \,, \tag{5.21}$$

with $\mu = \theta_0$ and $\sigma^2 = V_{00}$. The inverse of the covariance matrix is defined as

$$V_{ij}^{-1} = -E \left[\frac{\partial^2 \ln L}{\partial \theta_i \, \partial \theta_j} \right] \,, \tag{5.22}$$

with an assumed expectation value of μ'.

In the case of $\mu' = 0$ one can show that the pdf is given by

$$f(q_0|0) = \frac{1}{2}\delta(q_0) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_0}}e^{-q_0/2} \,, \tag{5.23}$$

which is a combination of a δ-function and a χ^2-distribution for one degree of freedom with a factor of $1/2$ each.

The cumulative distribution, given by the area under the pdf for the interval $[-\infty; q_0]$, is found to be

$$F(q_0|0) = \Phi(\sqrt{q_0}) \,. \tag{5.24}$$

Thus the p-value is given by

$$p_0 = 1 - F(q_0|0) \,. \tag{5.25}$$

Using eq. (5.3), the significance is obtained by the simple formula

$$Z_0 = \Phi^{-1}(1 - p_0) = \sqrt{q_0} \,. \tag{5.26}$$

6 Monte Carlo samples – Production, treatment & size

Machine learning algorithms, like BDT, are trained using Monte Carlo (MC) samples that simulate the outcome of the data taking process. These samples are produced (sec. 6.1), processed (sec. 6.2) and are then used for optimizing the event classification algorithm. Observables need to be found that show a high discrimination power between signal and background. The variables that are used as input for the BDTs are discussed in sec. 6.3. In sec. 6.4 the MC samples used for the different analyses carried out in the course of this thesis are described in more detail.

6.1 Production of the MC samples

For the production of the MC samples, firstly the collision events are calculated via mathematical models based on the SM. Thereafter a simulation of the detection process in the CMS detector is applied, including the detector geometry and its response to the collision events and detected particles. From that point on, the events are treated the same way as data. A reconstruction algorithm, considering the measured properties of the detected particles of an event, produces the final MC samples.

For the analysis, a preselection is applied that strongly reduces the number of background events. Only after this preselection, the MC samples are ready for constructing a BDT method and thereafter quantifying its performance. The BDT algorithm is trained using the MC samples to create an event classification method that is adapted to the problem and can thus be used for the final signal extraction when applied on data.

6.2 Preselection

To reach high performance in a BDT analysis, it is often of advantage to reduce the background before applying the BDT on the MC samples. Some machine learning algorithms perform better, if the ratio signal to background is higher, even though the number of signal events may be reduced. Nevertheless, the size of the MC sample should not be too small to ensure sufficient training potential. Only events that survive numerous preselection cuts reach the BDT classifier.

As explained in sec. 2.2.4 the analysis is only carried out for the VBF production mode and the channels $\mu\tau_h$ and $e\tau_h$. This leads to the requirement that only events with

one light lepton l (e or μ) and one τ decaying hadronically are included in the signal samples. Furthermore, these two particles must have opposite electric charge, as the Higgs boson is neutral. Due to VBF production two jets with large pseudorapidity gap $\Delta\eta_{j_1,j_2}$ contribute to the signature. Therefore at least two jets need to be present in the event with $\Delta\eta_{j_1,j_2} > 2.1$ between the two leading jets, denoting the two jets with highest transverse momentum p_T. Detected jets are only kept in an event if they fulfil requirements on several kinematic variables, namely $p_T > 30\,\text{GeV}$, $\eta < 4.7$ and a $\Delta R < 0.5$ between the jets and the objects τ_h, μ and e. η denotes the pseudorapidity (eq. 3.1) and

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \tag{6.1}$$

corresponds to a cone centred around the jet direction with ϕ being the polar angle. In addition, a central jet veto is applied. This veto removes events from the analysis containing a jet, which is not first or second leading jet and have a pseudorapidity η with either $\eta_1 < \eta < \eta_2$ or $\eta_2 < \eta < \eta_1$.

Similar to the jets also the other particles, namely τ_h, μ and e, have to fulfil several requirements on p_T, η, ΔR and I, denoting the isolation of the particle that quantifies the separation between the concerned particles and other particles.

Additionally, there are also requirements for combined quantities. For example, the transverse mass of the lepton-E_T^{miss}-system must have a value below $70\,\text{GeV}$ with

$$m_T = \sqrt{2 \cdot p_T(lep) \cdot p_T(E_T^{miss}) \cdot (1 - cos(\Delta\phi_{lep,E_T^{miss}}))}\,, \tag{6.2}$$

where E_T^{miss} denotes the missing transverse energy. The visible mass, meaning the mass of the system of the hadronic τ decay and the lepton, is required to be $m_{vis} > 40\,\text{GeV}$. Furthermore the baseline for the mass of the two-leading-jets-system is set to $m_{jj} > 500\,\text{GeV}$. Additional cuts are set on the isolation of the leptons, which is defined as

$$I(L) = \sum_{\text{charged}} p_T + \max\left(0, \sum_{\text{neutral}} p_T + \sum_{\gamma} p_T - \frac{1}{2}\sum_{\text{charged,PU}} p_T\right). \tag{6.3}$$

$\sum_{\text{charged}} p_T$ is the scalar sum of the p_T of all charged hadrons, electrons and muons that originate from the primary vertex and are located in a cone of size ΔR centred on the lepton direction. $\sum_{\text{neutral}} p_T$ and $\sum_{\gamma} p_T$ denote the sums the neutral hadrons and photons in the same cone, including neutral pileup particles. The pileup contribution of photons and neutral hadrons is estimated from the charged pileup particles ($\frac{1}{2}\sum_{charged,PU} p_T$) and subtracted from the total contribution of neutral particles. Events are rejected if the events contain b-tagged jets or additional leptons.

All preselection requirements are summarized in Tab. 6.1.

Tab. 6.1: Summary of all preselection requirements for τ_h, μ, e, jets and combined quantities.

Particle	Observable	Definition	Requirement
τ_h	p_T	transverse momentum of τ_h	$>20\,\text{GeV}$
	η	pseudorapidity of τ_h	<2.3
	I	isolation of τ_h	$<1\,\text{GeV}$
	ΔR	separation cone between τ_h & μ and τ_h & e	<0.5
μ	p_T	transverse momentum of μ	$>18\,\text{GeV}$
	η	pseudorapidity of μ	<2.1
	I	isolation of μ	$<0.15\,\text{GeV}$
	d_{xy}	distance between primary vertex and collision point in x-y-plane	$<0.045\,\text{cm}$
	d_z	distance between primary vertex and collision point in z-axis	$<0.2\,\text{cm}$
e	p_T	transverse momentum of e	$>23\,\text{GeV}$
	η	pseudorapidity of e	<2.1
	I	isolation of e	$<0.15\,\text{GeV}$
	ΔR	seperartion cone between μ & e	<0.15
jets	p_T	transverse momentum of the jet	$>30\,\text{GeV}$
	η	pseudorapidity of the jet	<4.7
	ΔR	separation cone between jets & other objects	<0.5
	$q_{lep} \cdot q_\tau$	product of the τ_h and lepton charge (opposite charge requirement)	<0
	no. jets	number of jets	≥ 2
	no. τ_h	number of τ_h	$=1$
	no. $e + \mu$	total number of e and μ	$=1$
	m_T	transverse mass of the lepton-E_T^{miss}-system	$<70\,\text{GeV}$
	m_{vis}	visible mass, mass of the μ-τ-system	$>40\,\text{GeV}$
	$\Delta\eta_{j_1,j_2}$	pseudorapidity gap between the two leading jets	>2.1
	CJV	central jet veto	✓
	bjetveto	veto on b-tagged jets	✓
	lepveto	veto on additional leptons	✓

6.3 Input variables for the analysis

It is essential to find variables with high discrimination power between signal and background events. For the 8 TeV and 13 TeV analysis almost the same input variables are calculated from the MC events with the only difference that for 8 TeV also the SVFit estimation of the transverse momentum of the τ-τ-system is included. Therefore, there are 13 variables taken into consideration in the 8 TeV analysis and only 12 for the 13 TeV analysis.

Tab. 6.2: List of the potential input variables for the BDT analysis that are taken into consideration.

Input variable	Definition
$m_{\tau\tau}^{SVFit}$	SVFit estimation of the mass of the τ-τ-system
$\Delta R_{lep,\tau}$	separation cone between lepton and τ_h
$\Delta \eta_{j_1,j_2}$	pseudorapidity gap between the two leading jets
m_{jj}	mass of the two-leading-jets-system
$\eta_{j_1} \cdot \eta_{j_2}$	product of the pseudorapidities of the two leading jets
p_T^{tot}	vector sum of the p_T of the lepton, τ_h, E_T^{miss} and the two jets
E_T^{miss}	missing transverse energy with respect to the visible decay products
m_T	transverse mass of the lepton-E_T^{miss}-system
η_{lep} centrality	pseudorapidity of the lepton with respect to the two leading jets
p_T^{sum}	scalar sum of the p_T of the lepton, τ_h, E_T^{miss} and the two jets
S	sphericity, quantity for the isotropy of the energy flow in the event
m_{vis}	visible mass of the lepton-τ-system
p_T^{SVFit}	SVFit estimation of the p_T of the two τ leptons

All investigated input variables are listed in Tab. 6.2 which contains also the η_{lep} centrality

$$C_{\eta_1,\eta_2}(\eta) = e^{-\frac{4}{(\eta_1-\eta_2)^2}\left(\eta-\frac{\eta_1+\eta_2}{2}\right)^2} \tag{6.4}$$

where η corresponds to the pseudorapidity of the lepton and η_1 and η_2 to the pseudorapidities of the two leading jets. All variables of the 13 TeV analysis for the $\mu\tau$ channel are shown in Fig. 6.1 and 6.2. Four MC samples are included in the histograms, the signal process $H \to \tau\tau$ and the background processes Drell-Yan production, $t\bar{t}$ and W+jets. As the different processes differ strongly in the number of expected events, each curve in the histogram is normalized to 1.
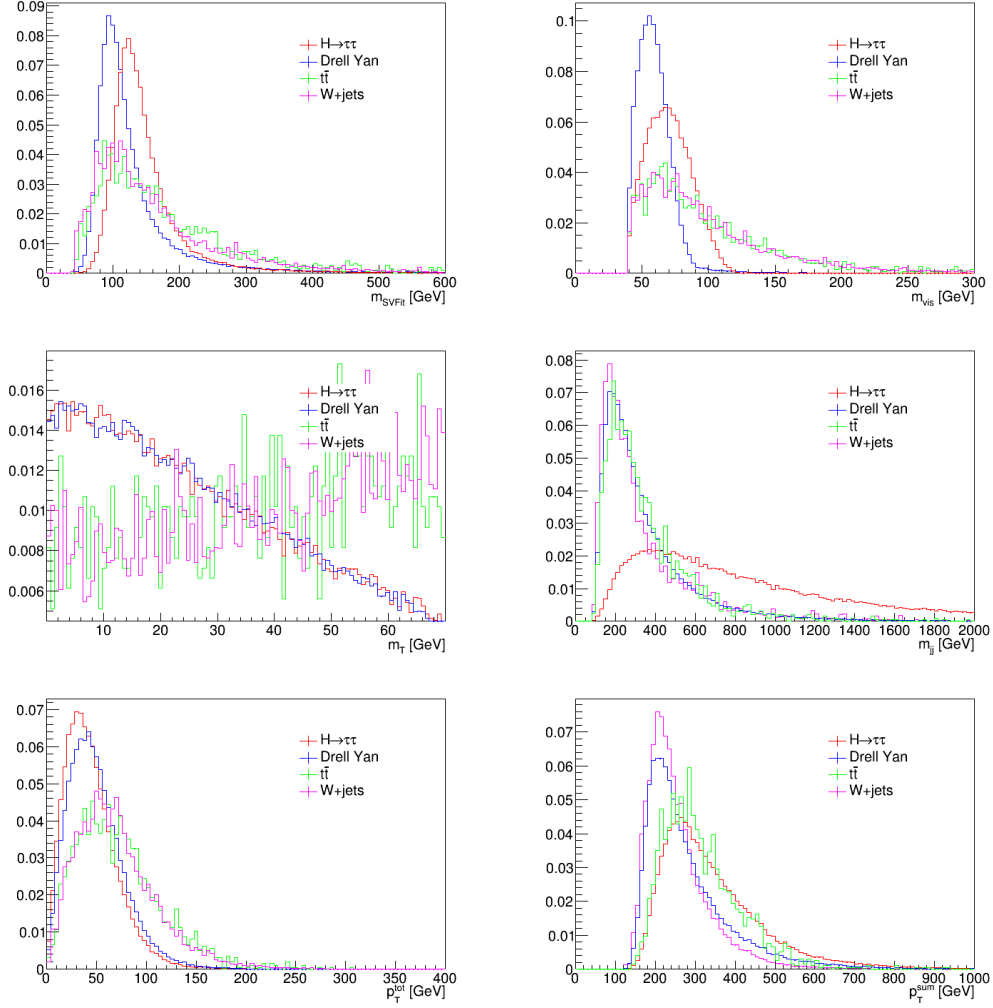
Fig. 6.1: Distribution of the variables $m_{\tau\tau}^{SVFit}$, m_{vis}, m_T, $m_{j,j}$, p_T^{tot} and p_T^{sum} for the 13 TeV MC samples with $H \to \tau\tau$, Drell-Yan, $t\bar{t}$ and W+jets in the $\mu\tau$ final state.
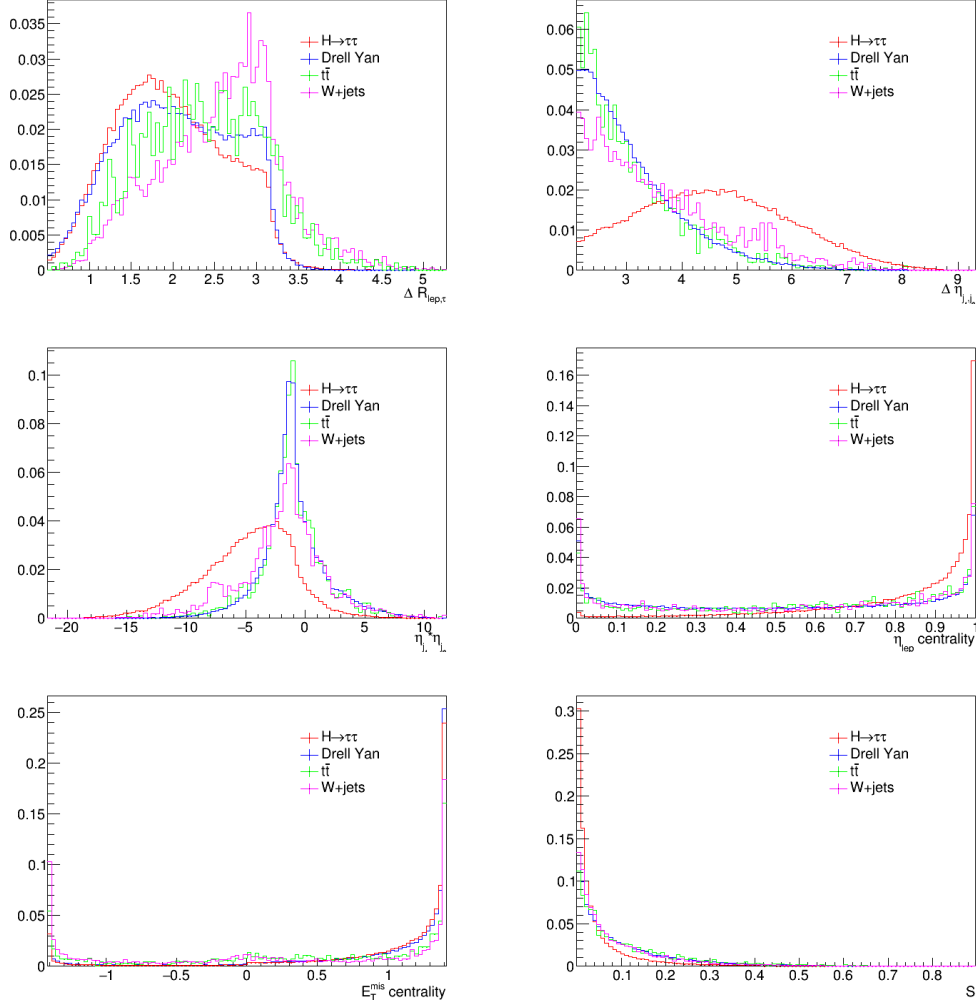
Fig. 6.2: Distribution of the variables $\Delta R_{lep,\tau}$, $\Delta \eta_{j_1,j_2}$, $\eta_{j_1} * \eta_{j_2}$, η_{lep} centrality, E_T^{miss} centrality and S for the 13 TeV MC samples $H \to \tau\tau$, Drell-Yan, $t\bar{t}$ and W+jets in the $\mu\tau$ final state.

6.4 Monte Carlo samples – Signal and background

The MC samples that are used in the analysis are produced following the procedure described in sec. 6.1 and then the preselection (see sec. 6.2) is applied that strongly reduces the size of the samples. Before the preselection is applied and the samples have reached there final size, each event needs to be reweighted in order to satisfy the physically expected yield. The expected number of events for each process naturally depends on the overall luminosity. Therefore, a weighting factor is calculated for all events according to

$$W_{lumi} = \frac{N_{exp}}{N_{MC}} \,, \tag{6.5}$$

with N_{MC} denoting the size of the MC samples before any selection and

$$N_{exp} = \mathcal{L} * \sigma \,, \tag{6.6}$$

where $\mathcal{L}$ is the luminosity and σ the production cross section.

As already stated in sec. 2.2.3 and 2.2.4 the events of the signal process originate from a Higgs boson produced via VBF. Only in the testing sample of the 8 TeV analysis several events are included generated through ggF in addition to the events produced via VBF. For 8 TeV only the $\mu\tau_h$ channel is analyzed as final state of the $H \to \tau\tau$, for 13 TeV also the $e\tau$ channel is added.

With respect to background processes, for the 8 TeV analysis only the Drell-Yan (DY) production is considered, whereas for the 13 TeV analysis three background sources are included, namely DY, $t\bar{t}$ and W+jets, which are described in more detail in sec. 2.2.4.

6.4.1 Analysis for 8 TeV

As for 8 TeV simply the $\mu\tau$ channel is considered, only two MC samples exist for the analysis, one for the VBF Higgs boson signal and one for the Drell-Yan background, which are split into half for training and testing purpose with a sample size shown in Tab. 6.3. The Drell-Yan sample listed in the table is split up into subsamples with respect to the number of jets in each event. Each subsample is reweighted with a different factor W_{lumi} to reach best alignment with physics. As additional signal events, a small number of Higgs events originating from ggF is added only for the testing procedure of the signal extraction algorithm, not for the training.

All 8 TeV samples are normalized to an integrated luminosity $\mathcal{L}$ of approximately $19.7 \, \text{fb}^{-1}$.

6.4.2 Analysis for 13 TeV

In contrast to the 8 TeV analysis, the samples for 13 TeV are normalized to an integrated luminosity of only $\mathcal{L} = 10 \, \text{fb}^{-1}$. Due to the higher cross section (XS) (see Fig. 2.4) and branching ratio (BR) at this energy a similar yield as for 8 TeV in terms of number

Tab. 6.3: Size of the MC samples for 8 TeV, final state $\mu\tau$, used for training ($N_{MC,train}$) and testing ($N_{MC,test}$) purpose, the total size $N_{MC,tot}$, the weighting factor W_{lumi} and the expected total number of events $N_{exp,tot}$ for $\mathcal{L} = 19.7 \, \mathrm{fb}^{-1}$. The table contains Higgs events from VBF and ggF production and the Drell-Yan background split according to the number of jets.

$$\mu\tau, \, 8 \, \mathrm{TeV}, \, \mathcal{L} = 19.7 \, \mathrm{fb}^{-1}$$

Process	$N_{MC,train}$	$N_{MC,test}$	$N_{MC,tot}$	W_{lumi}	$N_{exp,tot}$
H, VBF	5446	5449	10895	0.00180	20
H, ggF	-	669	669	0.02269	15
DY tot	5916	5899	11815	-	1423
DY 1jet	108	98	206	0.4385	90
DY 2jets	1731	1734	3465	0.1771	614
DY 3jets	2128	2122	4250	0.0988	420
DY 4jets	1949	1945	3894	0.0768	299

of signal events is reached. For 13 TeV not only the cross section and branching ratio change, but also the kinematic variables of the objects of the events have altered values and distributions.

Regarding the signal process, only the VBF production is included. In case of background processes, DY, $t\bar{t}$ and W+jets are taken into consideration. As the analysis is carried out for the $\mu\tau$ channel and later on also for the $e\tau$ final state, MC samples for both final states were produced.

In Tab. 6.4 and 6.5 the first four rows each denote the size of the inclusive MC samples with the preselection described in section 6.2. The samples for the signal and the DY background have reasonable size for BDT analysis and good statistics, whereas $t\bar{t}$ samples only provide 2-3 MC events for 1 expected event. It is not possible to use the W+jets sample for the training algorithm like BDT with only 48 events surviving the preselection. As workaround the preselection on the chargedIsoPtSum is loosened for the W+jets sample. Previously chargedIsoPtSum<1 was required, whereas now it is set to chargedIsoPtSum<1000. This modification is chosen as it is not strongly correlated with the input variables. It leads to a sample size for $\mu\tau$ of 1007 events (see Tab. 6.4, row 5), which is still very low in statistics as 1 MC event corresponds to 3 expected events.

In order to improve the training performance of the BDT with respect to the W+jets background, the inclusive MC sample (discussed above) is replaced with the the the h_T-binned sample. This h_T-binned sample is split into subsamples according to the h_T of the events with

$$h_T = \sum_{jets} p_T \, , \quad h_T > 100 \, . \tag{6.7}$$

The sample size of the four h_T-samples is shown in Tab. 6.4. Each sample covers a

Tab. 6.4: List of the total size of the MC samples $N_{MC, tot}$ for 13 TeV and the final state $\mu\tau$, the weighting factor W_{lumi} and the expected total number of events $N_{exp, tot}$ for $\mathcal{L} = 10\,\text{fb}^{-1}$. The table contains Higgs events from VBF production and the Drell-Yan, $t\bar{t}$ and W+jets background.

$\mu\tau$, $13\,\text{TeV}$, $\mathcal{L} = 10\,\text{fb}^{-1}$

Process	$N_{MC, tot}$	W_{lumi}	$N_{exp, tot}$
H, VBF	89764	0.000240	22
DY	78404	0.0123	964
$t\bar{t}$	1963	0.4160	817
W+jets	48	61.42	2948
W+jets_loose	1007	2.928	2948
W+jet_ht_tot	1644		2442
W+jets_ht1	385	4.2470	1635
W+jets_ht2	626	1.1750	736
W+jets_ht3	409	0.1474	60
W+jets_ht4	224	0.0505	11
W+jet_ht_loose_tot	78443		2442
W+jets_ht1_loose	5633	0.29030	1635
W+jets_ht2_loose	19471	0.03778	736
W+jets_ht3_loose	27244	0.00221	60
W+jets_ht4_loose	26095	0.00043	11
W+jets_ omb	1784		2948
W+jets_loose_comb	78583		2948

Tab. 6.5: List of the total size of the MC samples $N_{MC,tot}$ for 13 TeV and the final state $e\tau$, the weighting factor W_{lumi} and the expected total number of events $N_{exp,tot}$ for $\mathcal{L} = 10\,\text{fb}^{-1}$. The table contains Higgs events from VBF production and the Drell-Yan, $t\bar{t}$ and W+jets background.

$e\tau$, 13 TeV, $\mathcal{L} = 10\,\text{fb}^{-1}$

Process	$N_{MC,tot}$	W_{lumi}	$N_{exp,tot}$
H, VBF	52025	0.00239	12
DY	34351	0.00707	243
$t\bar{t}$	1110	0.41600	462
W+jets	30	61.48	1845
W+jets_loose	647	2.851	1845
W+jet_ht_tot	1080		1411
W+jets_ht1	192	4.3620	838
W+jets_ht2	446	1.1750	524
W+jets_ht3	283	0.1474	42
W+jets_ht4	159	0.0512	8
W+jet_ht_loose_tot	55070		1411
W+jets_ht1_loose	3396	0.24660	838
W+jets_ht2_loose	13317	0.03936	524
W+jets_ht3_loose	19408	0.00215	42
W+jets_ht4_loose	18949	0.00043	8
W+jets_comb	1163		1845
W+jets_loose_comb	55153		1845

different h_T range, namely 100-200 (ht1), 200-400 (ht2), 400-600 (ht3) and 600-inf (ht4). This means that the h_T range from 0 to 100 is not covered by these samples. In order not to neglect these events in the analysis, events with $h_T < 100$ are extracted from the loose inclusive sample and added to the h_T-binned samples. To fulfil the requirement of a total expected number of events of 2948 (1845) for the $\mu\tau$ ($e\tau$) channel the events added to the h_T-binned samples must be reweighted. The new weighting factor is given by

$$W_{lumi,new} = \frac{\Delta N_{exp}}{\Delta N_{MC}} \tag{6.8}$$

where ΔN_{exp} is the difference in expected events between the inclusive and the sum of all h_T-binned samples, which leads to $\Delta N_{exp} = 506\,(434)$. ΔN_{MC} is the number of events in inclusive sample with $h_T < 100$, namely $\Delta N_{MC} = 140\,(83)$. It must be kept in mind that for the h_T-combined sample events with original (h_T-binned) and loose (inclusive) preselection are combined.

To further increase the training performance of the BDT and lower the statistical uncertainties, it is of advantage to further enlarge the sample size of the W+jets background. For this reason the h_T-binned samples are also generated with loose preselection leading to a sample size also shown in Tab. 6.4 and 6.5 (W+jets_ht_loose_tot). These samples are then again complemented by events with $h_T < 100$ of the loose inclusive sample.

To summarize, two bigger samples of the W+jets background are generated. On the first one the original preselection is applied. It covers the whole h_T range with a sample size of 1784 (1163) events (Tab. 6.4 (6.5), W+jets_comb). The second one fulfils the loose preselection, leading to a MC sample containing 78583 (55153) events (Tab. 6.4 (6.5), W+jets_loose_comb). It is essential to keep in mind that these two samples are NOT independent from each other, as all events of the smaller sample also appear in the bigger one.

7 Multivariate analysis using the 8 TeV Monte Carlo sample

In sec. 4.4 the principles of the a boosted decision tree are introduced and the most important tuning parameters of the algorithm are explained.

There are two optimization procedures to obtain the best result for the signal extraction. These general optimization processes are explained below, but they are not restricted to the analysis of the 8 TeV samples. The optimization procedure for the 13 TeV analysis (see Chap. 8) follows the same pattern as shown in this chapter.

As stated in sec. 6.3, 13 input variables are tested. With this set of variables a rough optimization is carried out on the BDT parameters, which are introduced in sec. 4.4. For machine learning methods sometimes a smaller set of input variables leads to a higher FoM. Thus it can be of advantage, also in terms of computation time, to reduce the set of input variables to the most powerful ones. To find the best subset one optimization of the input variable set is performed. After this choice is made, a proper optimization of the BDT parameters is performed.

7.1 Optimization of BDT parameters

The first step of performing an optimization of the BDT parameters is choosing an initial parameter set by an educated guess. The goal is to perform a multi-dimensional optimization with mostly continuous parameters. This is broken down into 1-dimensional scans of all parameters, to get a feeling for the impact of the single parameters on the performance. The parameter set is subsequently adjusted under consideration of the gained information. Often, it is of advantage not to choose the global maximum as working point, especially if it is located in an unstable region with a high variation in the AMS values in the neighbourhood.

In Fig. 7.1 and 7.2 several 1-dimensional scans of the BDT parameters are displayed. The difference between training and testing set gives information on the magnitude of the overtraining. E.g. for MaxDepth a significant overtraining occurs starting from MaxDepth=3, with firstly a stagnation and then even a decrease of the performance of the algorithm. A similar observation is made for Shrinkage. The higher the Shrinkage the stronger the adaptation to the fluctuation of small samples. Another effect of fluctuations can be noticed in the plot for MinNodeSize. For low values the constructed trees have very small final nodes. Therefore it is much more probable that the trees are dominated by fluctuations. On the other hand, some BDT parameters affect the performance of

Fig. 7.1: FoM (AMS) as a function of different BDT parameters using the MC samples for the signal $H \to \tau\tau$ and the Drell-Yan background (red line: FoM for training samples, blue line: FoM for test samples), left: 8 TeV, right: 13 TeV.
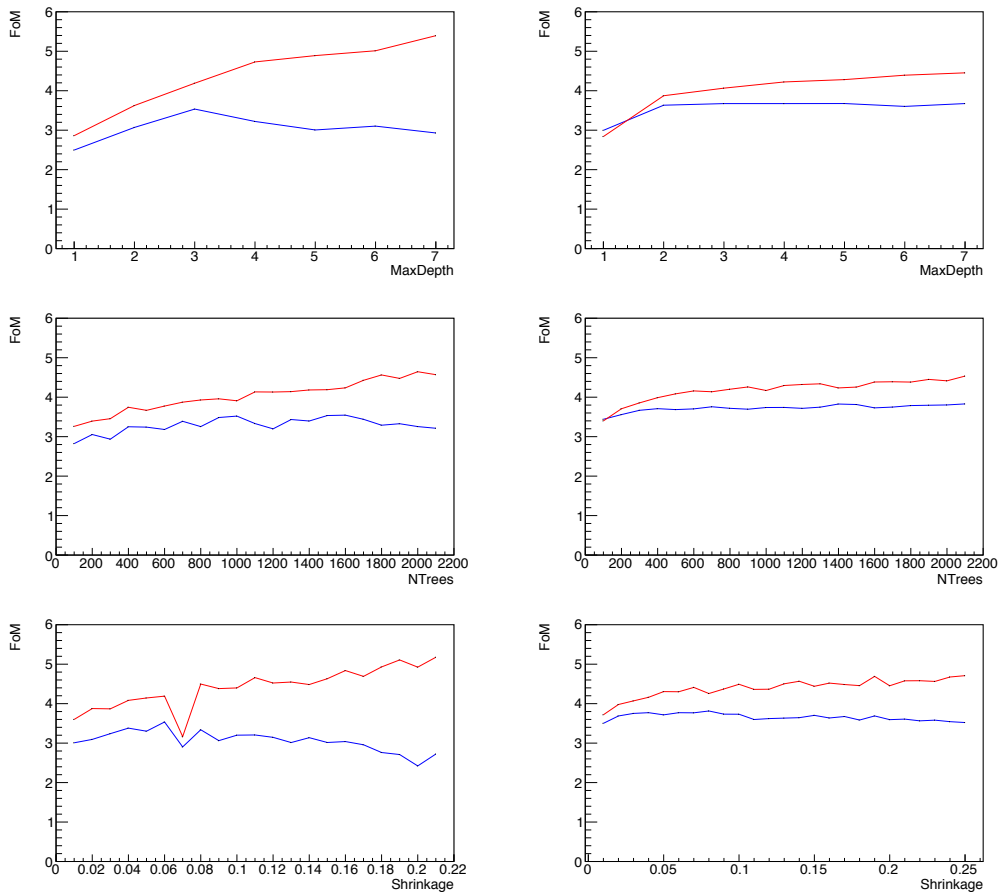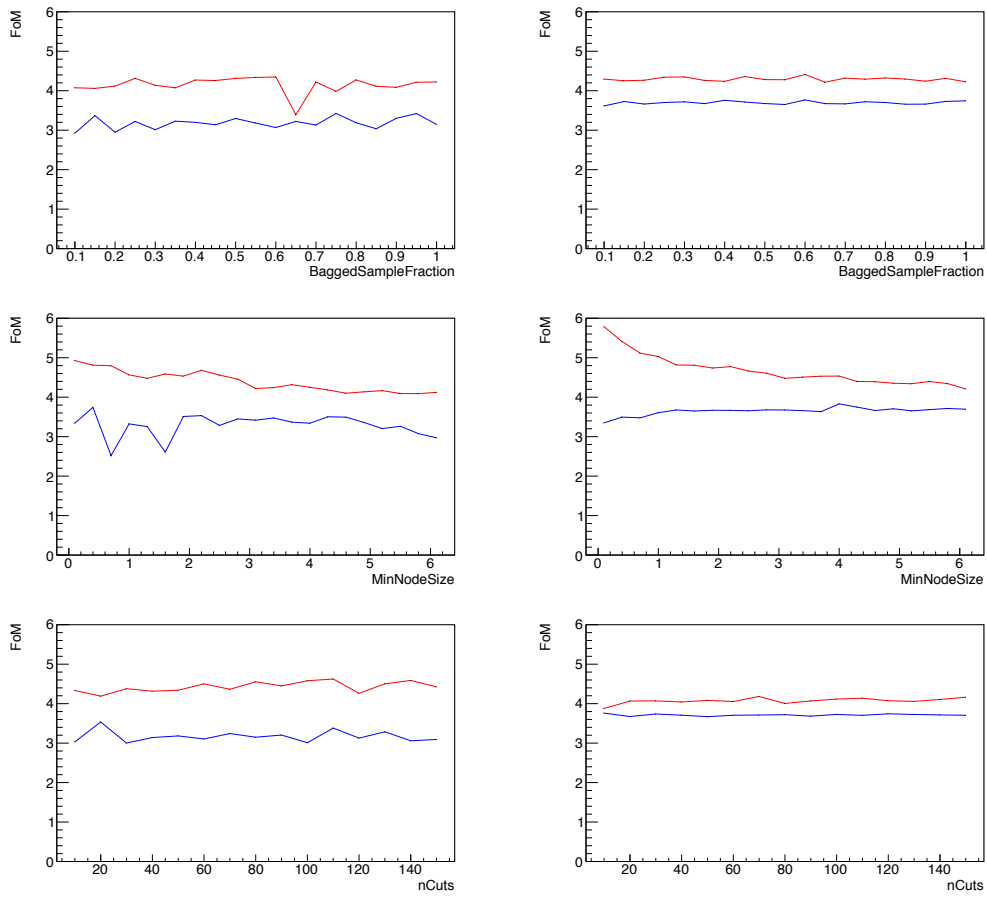
Fig. 7.2: FoM (AMS) as a function of different BDT parameters using the MC samples for the signal $H \rightarrow \tau\tau$ and the Drell-Yan background (red line: FoM for training samples, blue line: FoM for test samples), left: 8 TeV, right: 13 TeV.

the analysis only slightly, but do have a strong influence on the computation time, like nCuts. In order to save some computation time, nCuts is chosen relatively small, namely nCuts=20.

Although in this chapter only the BDT parameter optimization is discussed, the parameter scans shown in the plots are already performed with the fully optimized parameter and variable set, to show the dependency of the final BDT on each parameter. This fully optimized parameter and variable set is shown in Tab. 7.1 and 7.2.

In addition to the 1-dimensional scans, also 2-dimensional scans of the parameters are performed which show better the impact of correlations between the chosen parameters. Two examples are given in Fig. 7.3. The upper plot shows the 2D scan of MinNodeSize and NTrees, the second plot MinNodeSize and Shrinkage. In the graphics, stable regions can be identified for a MinNodeSize of approximately 4-5%, a number of trees (NTrees) of 1400-1600 and a Shrinkage of 0.05-0.08.

In the plot on top of Fig. 7.3 one can identify a region with overall low AMS on the left side of the plot. This region is characterized by a very low number of trees, which indicates that this low number of trees is not sufficient for a proper training of the BDT algorithm. Another distinct region, which shows strong fluctuations and low AMS, is located in the lower right edge. Also for the bottom plot in Fig. 7.3 a strongly fluctuating region with low AMS can be detected. This fluctuations are caused by two reasons, firstly a high learning rate (Shrinkage) that allows fluctuations in the MC samples to have a high impact on the BDT and secondly a low MinNodeSize, where outliers have more influence on the analysis.

Similar to the 1D scans, also the 2D scans illustrated are already performed with the optimized variable and parameter set to show the final dependencies and AMS values.

The final parameter set that is chosen for the 8 TeV analysis with the $H \rightarrow \tau\tau$ signal and Drell-Yan background is summarized in Tab. 7.1 .

Tab. 7.1: Values of BDT parameters after optimization for the 8 TeV analysis.

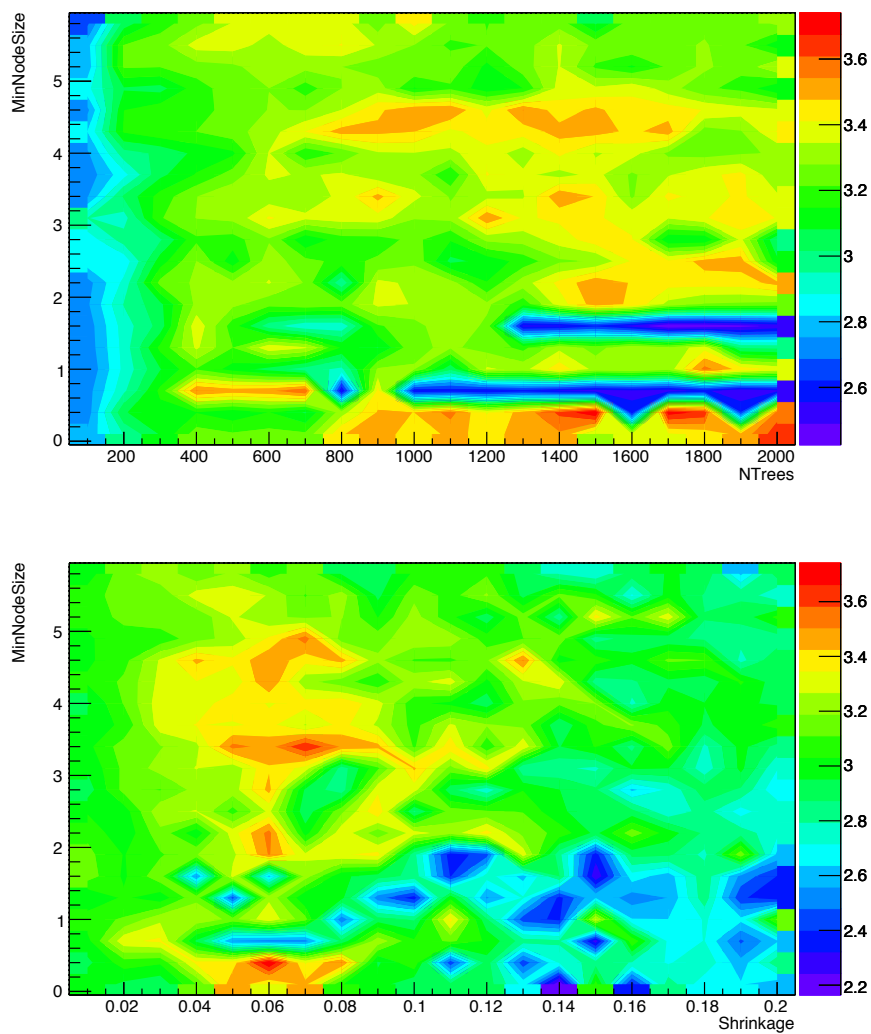| BDT parameter	value
NTrees	1500
Shrinkage	0.06
MinNodeSize	4.4%
MaxDepth	3
nCuts	20
UseBaggedSample	✓
BaggedSampleFraction	0.75
BoostType	GradBoost

Fig. 7.3: 2D scans of BDT parameters, top: Figure of Merit (AMS) as a function MinNode-Size and NTrees, bottom: Figure of Merit (AMS) as a function of MinNodeSize and Shrinkage.

7.2 Optimization of the input variable set

As stated earlier 13 input variables are available for the 8 TeV analysis, which are listed in Tab. 6.2 . To choose the optimal variable subset that leads to the highest performance of the signal extraction a 13-dimensional optimization problem needs to solved. Using the brute force method, which corresponds to testing every single possibility, is not feasible for such a high dimensional problem. Therefore, a different approach is chosen for finding the best variable set.

Starting with the whole set, the set of variables is reduced in size, always eliminating the least important variable. This means that starting with 13 variables, each variable is excluded once and the obtained BDT is evaluated. The subset that reaches the highest AMS is then chosen as new starting point. From there again every variable is excluded once and so on, until only a single variable is used as input. In the end the subset of input variables that reaches the highest AMS overall is taken as the optimal input variable set.

The best performance is obtained not with the full set of input variables, but with a subset of 8 variables. The performance even decreases using more input variables. Nevertheless, if the input set is reduced even further, the AMS decreases as well, as not enough information is provided any more for the classification of the events. In Fig. 7.4 the dependency of the AMS on the number of input variables is shown. The AMS steadily increases with the number of input variables until a subset of 8 input variables. Adding more variables, the performance decreases again. Therefore, a subset of 8 input variables is chosen for the analysis, including the variables listed in Tab. 7.2. In the table the variables are sorted by their importance on the performance of the BDT. The most important variable is the one which is eliminated last.
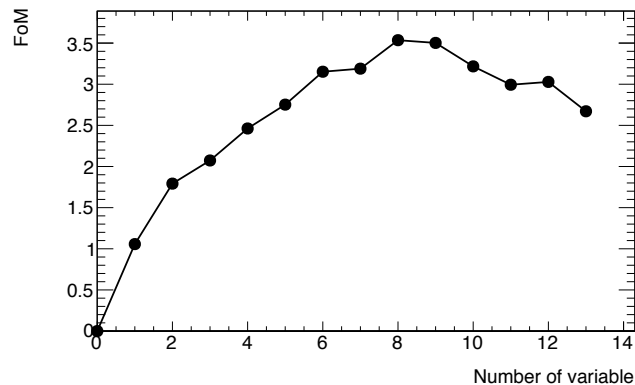

Fig. 7.4: Figure of Merit (AMS) as a function of the number of input variables in the BDT algorithm for 8 TeV using the BDT parameter values stated in Tab. 7.1.

Tab. 7.2: Subset of 8 input variables chosen for the 8 TeV analysis after optimization, sorted by their importance for the performance of the BDT.

Rank	Input variable
1.	m_{SVFit}
2.	$\Delta\eta_{j_1,j_2}$
3.	p_T^{sum}
4.	p_T^{tot}
5.	m_{vis}
6.	E_T^{miss} centrality
7.	$\Delta R_{lep,\tau}$
8.	η_{lep} centrality

7.3 Results of the optimized BDT analysis

After a full optimization of the BDT parameters and input variable set, the performance of the BDT method is evaluated. The classifier reaches a figure of merit (FoM), in this case an approximate median significance, of AMS=3.5. Fig. 7.5 shows the distribution of signal and background events after the event classification using the optimized BDT. The right side of the histogram corresponds to the signal region, the left side to the background region. The histogram shows the BDT score after the smart rebinning (see sec. 5.1), which causes the smooth increase in background from right to left.

7.4 Comparison of the BDT analysis with a cut-based analysis for 8 TeV

Until now the optimization and comparison between different classifiers only has been done for BDTs generated with different parameters and input variables. To get a feeling for the power of BDTs, it is useful to compare the performance of the constructed BDT with a different signal extraction method. For this comparison a cut-based analysis is chosen as this method was also used in earlier analyses [20]. The cut-based analysis throughout this thesis is optimized by the use of the TMVA package [33].

For the cut-based method, all surviving events distributed according to the SVFit mass estimation of the $\tau\tau$-system $m_{\tau\tau}^{SVFit}$ are used to evaluate the performance.

The histograms of the two methods that are used for evaluation are shown in Fig. 7.5 (BDT) and Fig. 7.6 (cut-based). The AMS that is reached after the full optimization is given by AMS=3.5 for the BDT and AMS=2.1 for the cut-based analysis. This shows a significant improvement of the BDT method compared to the cut-based analysis.

For extracting these results, several approximations are made here. First of all, only the Drell-Yan background $Z \to \tau_h\tau_h$ was included in the analysis. The AMS, which is used for quantification, only provides a simplified statistical treatment. In terms of sys-

tematics, only normalization uncertainties are included, but e.g. no shape uncertainties. For an adequate evaluation of the significance, the CMS CombinedLimit framework (see Chap. 9) can be implemented.

As these approximations are done, the comparison is only qualitative. BDT does give better results in this case, but it is not clear by how much its performance will exceed the cut-based analysis when a full analysis is performed.
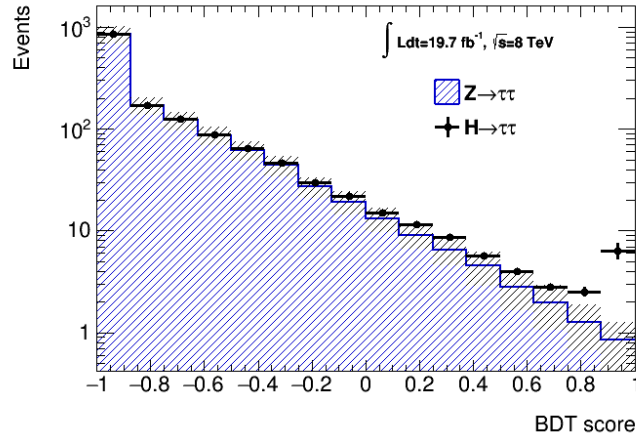
Fig. 7.5: Histogram showing the distribution of the BDT score for the signal process $H \rightarrow \tau\tau$ stacked on top of the Drell-Yan background process $Z \rightarrow \tau\tau$ in the final state $\mu\tau_h$ for 8 TeV normalized to a luminosity of $\mathcal{L} = 19.7\,\mathrm{fb}^{-1}$.

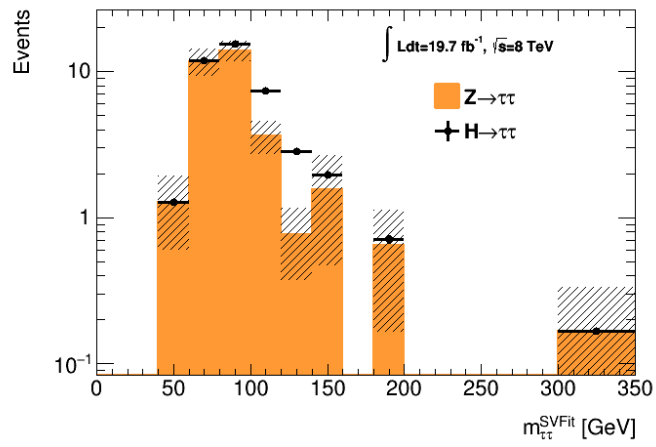Fig. 7.6: Histogram showing the $m_{\tau\tau}^{SVFit}$ distribution for the events that survived the cut-based analysis including the signal process $H \rightarrow \tau\tau$ stacked on top of the Drell-Yan background process $Z \rightarrow \tau\tau$ in the final state $\mu\tau_h$ for 8 TeV normalized to a luminosity of $\mathcal{L} = 19.7\,\text{fb}^{-1}$.

8 Multivariate analysis using the 13 TeV Monte Carlo sample

The main purpose of this thesis is to find the best BDT algorithm for the event classification and signal extraction of the $H \to \tau\tau$ decay in the final states $\mu\tau_h$ and $e\tau_h$ at 13 TeV under consideration of three backgrounds, Drell-Yan production, $t\bar{t}$ and W+jets. The BDT algorithm is optimized with respect to the BDT parameters and input variables using the same procedure as discussed in sec. 7.1 and 7.2.

As a first step, only the $H \to \tau\tau$ signal process and the Drell-Yan background in the $\mu\tau_h$ final state are investigated, which means that the processes considered are the same as in the analysis for 8 TeV. The initial set of optimized parameters and input variables, is the one obtained in the previous chapter. The biggest difference between the analysis for 8 TeV and 13 TeV consists in the different size of the MC samples available. As illustrated in Tab. 6.3, for the 8 TeV MC samples each process has a sample size of approximately 11000-12000 events, whereas for 13 TeV the samples contain 90000 and 78000 events for the signal and the Drell-Yan process in the $\mu\tau_h$ channel. The impact of the sample size on the BDT analysis can be studied.

The 8 TeV and 13 TeV MC samples are normalized to different luminosities, namely $\mathcal{L}_{8\,TeV} = 19.7\,\mathrm{fb}^{-1}$ and $\mathcal{L}_{13\,TeV} = 10\,\mathrm{fb}^{-1}$. The input variables have different values and the shapes of their distributions have changed. The performance that is reached in both analysis is not directly comparable.

Later on two additional backgrounds are included in the analysis, namely the $t\bar{t}$ and the W+jets background. As these backgrounds have different properties compared to the signal and Drell-Yan process, a new optimization of the BDT with respect to the parameters and the chosen input variable subset is performed.

Lastly the $e\tau_h$ final state is analysed. Two approaches of combining the two analyses of $\mu\tau_h$ and $e\tau_h$ are introduced. They are further discussed in chapter 9.

8.1 Optimization of the BDT with respect to the signal and DY processes ($\mu\tau_h$ channel)

Here, only the Drell-Yan background is included in the event classification method. For both the signal process $H \to \tau\tau$ and the DY background the final state $\mu\tau_h$ is considered.

To construct the BDT, the algorithm is optimized using the same procedure as in the previous chapter. The procedures of tuning the BDT parameters and choosing the most

powerful subsample of input variables is performed iteratively. The best parameters and variables obtained are summarized in Tab. 8.1 and 8.2.

When compared to the 8 TeV analysis, the parameters only show slight variations. The values of NTrees, MaxDepth, nCuts and the boosting type remain the same. The BaggedSampleFraction changes from 0.75 to 0.5 and the modifications in Shrinkage (0.06→0.07) and MinNodeSize (4.4%→4.0%) are minor. Regarding the input variables, the importance of single variable changes but the variable set stays the same, only $\Delta\eta_{j_1,j_2}$ is replaced by m_{jj}.

In Fig. 8.1 the reduction process of the variable set is illustrated. In contrast to 8 TeV, no obvious maximum emerges. As the 8 TeV analysis uses smaller MC samples, this maximum probably emerges as a consequence of overtraining, which is higher for a larger set of input variables. For 13 TeV, starting from one variable the FoM initially increases until an amount of 8 variables. From there on, the FoM stays stable and does not vary strongly for additional input variables. The subset with the 8 most powerful variables is chosen for the analysis.

For the optimized BDT an AMS of 3.7 is reached for the testing samples. The BDT score distribution that is used for calculating the AMS is given in Fig. 8.2.

Tab. 8.1: Values of BDT parameters for the 13 TeV analysis with final state $\mu\tau_h$ only including the Drell-Yan background.

BDT parameter	value
NTrees	1500
Shrinkage	0.07
MinNodeSize	4.0%
MaxDepth	3
nCuts	20
UseBaggedSample	✓
BaggedSampleFraction	0.5
BoostType	GradBoost

8.2 Comparison of BDT to a cut-based analysis and other machine learning approaches

The BDT analysis for the $\mu\tau_h$ channel including the DY background carried out in the previous section gives an AMS of 3.7. A cut-based analysis is performed leading to AMS=1.92 with the event distribution shown in Fig. 8.3.

Parallel to this thesis two other MVA methods are tested on the same MC samples, namely NeuroBayes and deep neural networks. The NeuroBayes analysis is carried out by the use of the TMVA package [33]. After optimization an AMS of 3.1 [38] is obtained. Deep neural networks reach an AMS of 2.15 [17].

Tab. 8.2: Subset of 8 input variables chosen for the 13 TeV analysis with final state $\mu\tau_h$ including DY background, sorted by their importance for the performance of the BDT.

Rank	Input variable
1.	m_{jj}
2.	m_{vis}
3.	η_{lep} centrality
4.	$m_{\tau\tau}^{SVFit}$
5.	p_T^{tot}
6.	p_T^{sum}
7.	$\Delta R_{lep,\tau}$
8.	E_T^{miss} centrality
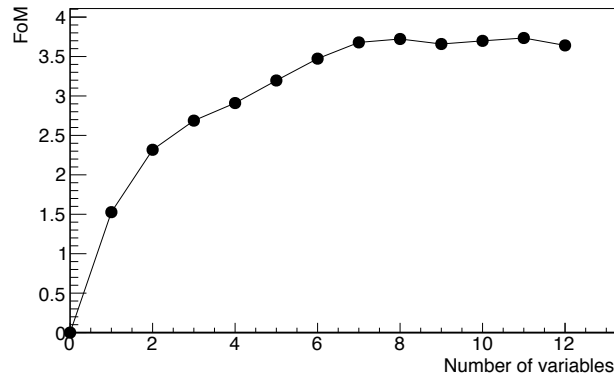
Fig. 8.1: Figure of Merit (AMS) as a function of the number of input variables in the BDT algorithm for 13 TeV using the BDT parameter values stated in Tab. 8.1. Included processes: signal, DY background; final state: $\mu\tau_h$.
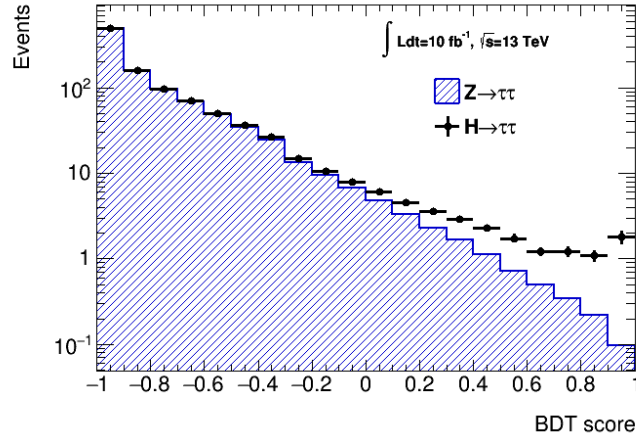
Fig. 8.2: Histogram showing the distribution of the BDT score for the signal process $H \to \tau\tau$ stacked on top of the Drell-Yan background process $Z \to \tau\tau$ with final state $\mu\tau_h$ for 13 TeV normalized to a luminosity of $\mathcal{L} = 10\,\mathrm{fb}^{-1}$.
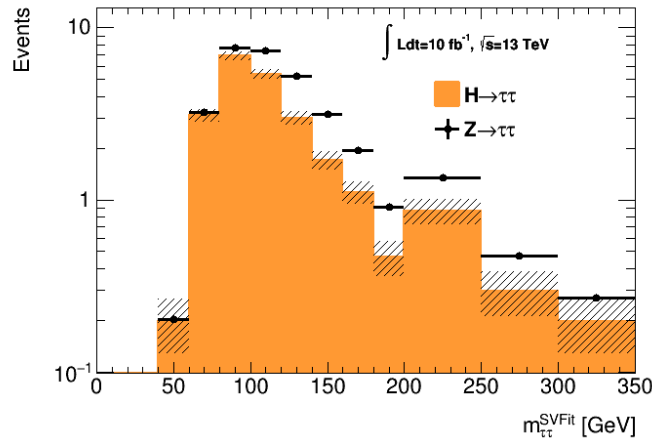
Fig. 8.3: Histogram showing the surviving events after a cut-based analysis with distribution variable $m_{\tau\tau}^{SVFit}$ for the signal process $H \to \tau\tau$ stacked on top the Drell-Yan background process $Z \to \tau\tau$ with final state $\mu\tau_h$ for 13 TeV normalized to a luminosity of $\mathcal{L} = 10\,\mathrm{fb}^{-1}$.

Summarizing, for the $\mu\tau_h$ analysis including the DY background the BDT analysis performs significantly better than the other approaches. The deep neural network only gives a slightly better result than the cut-based analysis. The reason for this is most probably a too small sample size of the MC training samples.

8.3 Impact of the MC sample size on the BDT performance

As the MC samples for the signal and DY processes for 8 TeV and 13 TeV significantly vary in size for the two analyses, the 1-dimensional scans performed for the optimization purpose can also be used for studying the impact of the sample size on the choice of the BDT parameters. The 1D scans are shown in Fig. 7.1 and 7.2.

One can extract two main consequences of bigger samples. Firstly, the variations between obtained FoMs for similar parameter sets are reduced, which corresponds to a reduced influence of fluctuations for bigger samples. Secondly, the red line is closer to the blue one, which equals to less overtraining. The bigger the samples, the less the probability for the machine learning algorithm to overtrain on outliers instead of adapting to global characteristics.

8.4 Performance of the BDT considering the DY, $t\bar{t}$ and W+jets background for the $\mu\tau_h$ channel

In order to improve the analysis with respect to a more complete background treatment, the $t\bar{t}$ and W+jets background is added. The size of the MC samples is shown in Tab. 6.4. The $t\bar{t}$ and especially the W+jets background samples are significantly smaller than the others as discussed in detail in sec. 6.4.2. For this reason, the preselection is loosened and a special sample is used for W+jets.

As a first attempt the new backgrounds are included in the testing samples, whereas the BDT algorithm itself remains unchanged: The used BDT is only trained on the DY background. This does not lead to a satisfying outcome. The observed AMS drops from 3.7 (only DY) to 0.6. The event distribution is shown in Fig. 8.4. The W+jet background clearly contributes strongly to the background in the signal region, but also $t\bar{t}$ has a high impact.

A possible way out is the generation of a new BDT method which is trained on all three backgrounds. In order not to distort the quantification of the BDT performance by events that only survive the analysis due to the loosened preselection, the W+jets sample with original preselection is used for testing purpose. On the other hand, for training it is essential to use MC samples as big as possible to prevent overtraining, which motivates the use of the loose W+jets sample there.

The new BDT method undergoes the optimization process leading to the input variable set and BDT parameters shown in Table 8.3 and 8.4. Compared to the variable subset used for the BDT analysis in sec. 8.1, the two variables m_{jj} and m_{vis} are replaced

by $\Delta\eta_{j_1,j_2}$ and m_T, the others stay the same. Having a look at the input variable distributions in Fig. 6.1 and 6.2, at least for the variable m_{vis} it is obvious that it looses importance, as the new backgrounds do not show an as distinct discrimination to the signal as the DY process. m_T gains separation power, as the new backgrounds show exactly the opposite distribution compared to the old processes. The most significant change in the BDT parameters concerns Shrinkage, which now has value of 0.17, compared to 0.07 before. The other variables only experience minor modifications. NTrees increases from 1300 to 1500, MinNodeSize from 3.5% to 4.0%, nCuts form 20 to 40 and BaggedSampleFraction for 0.4 to 0.5.
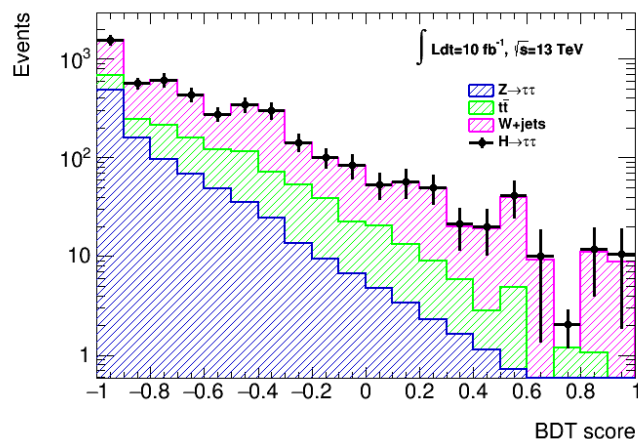
Fig. 8.4: Distribution of the BDT score for the signal process $H \to \tau\tau$ stacked on top of the 3 backgrounds DY ($Z \to \tau\tau$), $t\bar{t}$ and W+jets, with the final state $\mu\tau_h$ for 13 TeV normalized to a luminosity of $\mathcal{L} = 10\,\mathrm{fb}^{-1}$. The BDT is only trained on DY.

Testing the BDT trained on all three background samples gives an AMS of 2.6 (for the BDT score distribution, see Fig. 8.5), which is still significantly lower than without the $t\bar{t}$ and W+jets backgrounds. Nevertheless, it shows a significant improvement compared to the result obtained by the BDT only trained on DY. The still relatively low statistics of the new backgrounds strongly affect the uncertainties in the event distribution leading to bigger uncertainties in the region with more W+jets background events, as visible in the plot. To compare with a different event classifier the cut-based analysis is extended to 3 backgrounds giving an AMS of 1.05. With respect to the cut-based analysis performed only with DY (see Fig. 8.3), the optimal event selection is now much tighter. The distribution of the events is illustrated in Fig. 8.6.

8.5 BDT analysis with a different final discriminant

So far the BDT score obtained during the BDT classification is used as final discriminant. It is conceivable that a better performance can be obtained by using a different

Tab. 8.3: Subset of 8 input variables chosen for the 13 TeV analysis, final state $\mu\tau_h$, including DY, $t\bar{t}$ and W+jets background, sorted by their importance.

Rank	Input variable
1.	$\Delta\eta_{j_1,j_2}$
2.	p_T^{sum}
3.	$m_{\tau\tau}^{SVFit}$
4.	p_T^{tot}
5.	η_{lep} centrality
6.	E_T^{miss} centrality
7.	$\Delta R_{lep,\tau}$
8.	m_T

Tab. 8.4: Values of BDT parameters for the 13 TeV analysis, final state $\mu\tau_h$, including the Drell-Yan, $t\bar{t}$ and W+jets background.

BDT parameter	value
NTrees	1300
Shrinkage	0.17
MinNodeSize	3.5 %
MaxDepth	3
nCuts	40
UseBaggedSample	✓
BaggedSampleFraction	0.4
BoostType	GradBoost
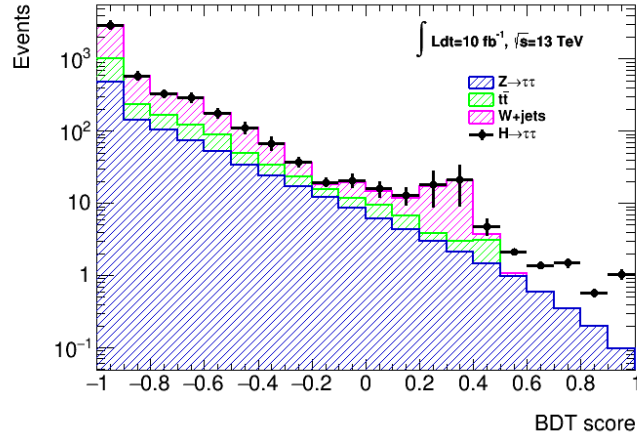
Fig. 8.5: BDT score distribution for the signal process $H \to \tau\tau$ stacked on top of the Drell-Yan ($Z \to \tau\tau$), $t\bar{t}$ and W+jets background processes for 13 TeV in the final state $\mu\tau_h$ normalized to a luminosity of $\mathcal{L} = 10\,\mathrm{fb}^{-1}$.
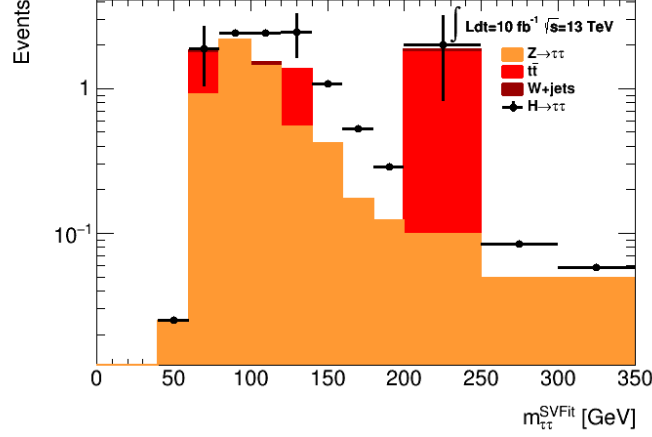
Fig. 8.6: Histogram showing the surviving events after a cut-based analysis with $m_{\tau\tau}^{SVFit}$ as distribution variable for the signal process $H \rightarrow \tau\tau$ stacked on top of the Drell-Yan ($Z \rightarrow \tau\tau$), $t\bar{t}$ and W+jets background processes with final state $\mu\tau_h$ normalized to a luminosity of $\mathcal{L} = 10\,\text{fb}^{-1}$.

distribution.

Like for the cut-based analysis, only a few events are used that survive a cut on the BDT score. Several variables are tested as discriminants, like $m_{\tau\tau}^{SVFit}$ and m_{vis}, and scans are performed to find the optimal cut on the BDT score. With respect to the other discriminants, m_{vis} gives the best results with the cut BDTscore>0.95. The cut value applies to the original BDT score before the smart rebinning is performed. The surviving events are then distributed as a function of their visible mass m_{vis}. This variable is excluded from the input variable set of the BDT. The method gives AMS=2.13, which is lower than for the standard BDT evaluation method used before that gave AMS=2.61. Nevertheless, it is still significantly higher than the performance of the cut-based analysis with an AMS of 1.45. The m_{vis} distribution is shown in Fig. 8.7.

In order to compare this method to the one used for the cut-based analysis, a similar number of surviving signal events is required, namely $N_{exp,sig} = 3.35$. Under this condition the cut on the BDT score is set to BDTscore>0.9, leading to AMS=1.58, which is also higher than for the cut-based analysis.

8.6 BDT analysis for the $e\tau_h$ channel

Firstly the analysis of the $e\tau_h$ final state including the three backgrounds Drell-Yan, $t\bar{t}$ and W+jets is performed using the BDT generated for the $\mu\tau_h$ channel. As this analysis did not give a satisfying result - the AMS only reaches a value of 1.7 - a new BDT is trained. The optimized values are listed in Tab. 8.5 and 8.6.

The strongest variable in the $\mu\tau_h$ channel, $\Delta\eta_{j_1,j_2}$, does not feature in the optimized
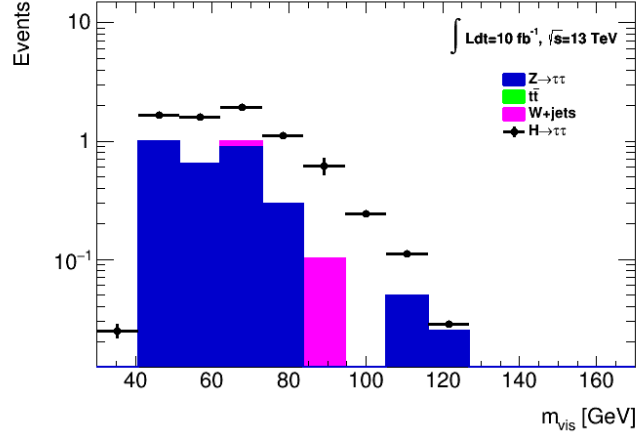
Fig. 8.7: m_{vis} distribution after a cut on the BDT score. Processes included in the analysis: signal process $H \to \tau\tau$, DY $Z \to \tau\tau$, $t\bar{t}$ and W+jets background, in the final state $\mu\tau_h$ normalized to a luminosity of $\mathcal{L} = 10\,\mathrm{fb}^{-1}$.

variable set of $e\tau_h$, whereas the strongest variable for $e\tau_h$, m_jj, is removed from the $\mu\tau_h$ analysis. The other variables show altered importance for the analysis as well and the variable η_{lep} centrality is replaced with p_T^{tot}. The BDT parameters do not change strongly. The most significant changes concern NTrees (1300→1600) and Shrinkage, which is reduced from 0.17 to 0.13. BaggedSampleFraction is increased from 0.4 to 0.6, the other parameters remain unchanged.

The BDT determined by the these parameter values and input variables gives AMS=2.0. The corresponding event distribution is shown in Fig. 8.8.

Tab. 8.5: Subset of 8 input variables chosen for the 13 TeV analysis in the $e\tau_h$ final state including DY, $t\bar{t}$ and W+jets background, sorted by their importance.

Rank	Input variable
1.	m_{jj}
2.	$m_{\tau\tau}^{SVFit}$
3.	η_{lep} centrality
4.	p_T^{tot}
5.	m_{vis}
6.	m_T
7.	p_T^{sum}
8.	$\Delta R_{lep,\tau}$

Tab. 8.6: Values of the BDT parameters for the 13 TeV analysis in the $e\tau_h$ final state including the Drell-Yan, $t\bar{t}$ and W+jets background.

| BDT parameter	value
NTrees	1600
Shrinkage	0.13
MinNodeSize	3.5 %
MaxDepth	3
nCuts	40
UseBaggedSample	✓
BaggedSampleFraction	0.6
BoostType	GradBoost
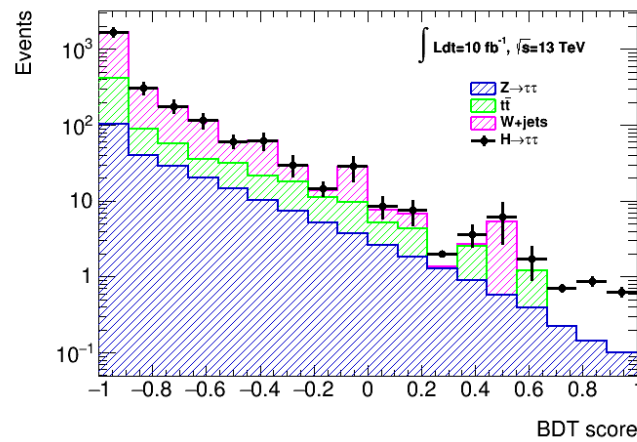
Fig. 8.8: Distribution of the BDT score for the signal process $H \to \tau\tau$ stacked on top of the Drell-Yan ($Z \to \tau\tau$), $t\bar{t}$ and W+jets background processes for 13 TeV in the final state $e\tau_h$ normalized to a luminosity of $\mathcal{L} = 10\,\mathrm{fb}^{-1}$.

8.7 Combination of the $\mu\tau_h$ and $e\tau_h$ final states

As the event classifiers are generated for both the $\mu\tau_h$ and the $e\tau_h$ channel in the course of the previous chapters, the next logical step is to combine the results obtained from these analyses. The aim is to enhance the performance of the signal extraction by this combination, as the information of both channels is used. For the combined analysis two different approaches are studied.

For the first method, the two channels are analysed separately using the BDTs generated in the previous chapters. The BDT score distributions obtained are narrowly binned. Subsequently, the content of the bins of the two histograms is summed up for each bin separately. The resulting BDT score distribution is then handled in the same way as the previous distributions in order to calculate the AMS. Performing this procedure an AMS of 2.9 is reached using the event distribution containing MC events from both the $\mu\tau_h$ and the $e\tau_h$ final state. To justify using this method, it is instructive to check that the distributions of the $\mu\tau_h$ and $e\tau_h$ channel do not differ much and the separation power is preserved. In Fig. 8.9 the distributions for the different processes are shown before the smart rebinning is performed. Only small differences are visible in the distributions.

The second method follows the idea of a new BDT trained on the whole range of events available for both final states and considering these two channel as one big new channel. For this purpose, the BDT is reoptimized leading to the input variables and BDT parameter values shown in Tab. 8.7 and 8.8. The optimized BDT gives an AMS of 2.9 with the corresponding event distribution shown in Fig. 8.10.

The set of the combined analysis is very similar to the variable set of the $\mu\tau_h$ analysis, except that in the combined analysis 7 variables are used instead of 8. The variable that is removed form the set is $\Delta R_{lep,\tau}$. Compared to the $e\tau_h$ channel, E_T^{miss} centrality is added, whereas m_{jj} and m_{vis} are removed from the set. In terms of BDT parameters, the same values are used as for the $e\tau_h$ analysis.

Tab. 8.7: Subset of 7 input variables chosen for the 13 TeV analysis for the combined analysis of the $\mu\tau_h$ and $e\tau_h$ final state including DY, $t\bar{t}$ and W+jets background, sorted by their importance.

Rank	Input variable
1.	$\Delta\eta_{jj}$
2.	p_T^{sum}
3.	$m_{\tau\tau}^{SVFit}$
4.	E_T^{miss} centrality
5.	p_T^{tot}
6.	η_{lep} centrality
7.	m_T
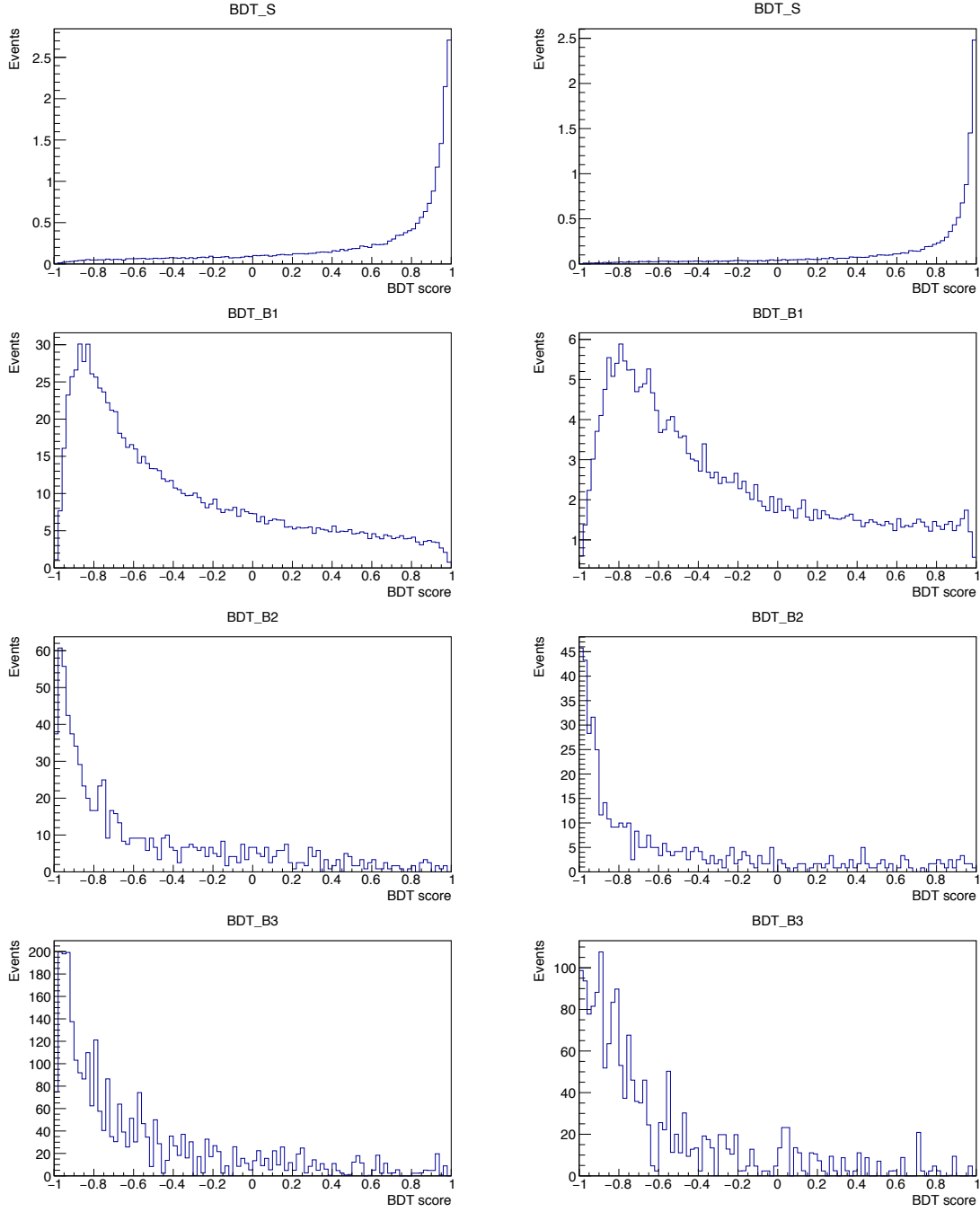
Fig. 8.9: BDT score distribution for the signal process (BDT_S), the DY (BDT_B1), $t\bar{t}$ (BDT_B2) and W+jets (BDT_B3) background for the $\mu\tau_h$ (left) and $e\tau_h$ (right) final states for the 13 TeV MC samples.

Tab. 8.8: Values of BDT parameters for the 13 TeV analysis for the combined analysis of the $\mu\tau_h$ and $e\tau_h$ final states including the Drell-Yan, $t\bar{t}$ and W+jets background.

BDT parameter	value
NTrees	1600
Shrinkage	0.13
MinNodeSize	3.5 %
MaxDepth	3
nCuts	40
UseBaggedSample	✓
BaggedSampleFraction	0.6
BoostType	GradBoost
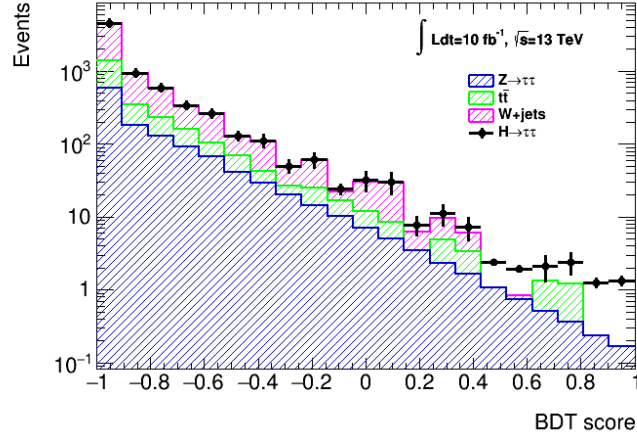
Fig. 8.10: Distribution of the BDT score for all processes normalized to a luminosity of $\mathcal{L} = 10\,\mathrm{fb}^{-1}$. The BDT is trained on both final states, $\mu\tau_h$ and $e\tau_h$.

Tab. 8.9: Summary of BDT parameters and input variables for all analyses carried out for 13 TeV.

13 TeV	$\mu\tau_h$		$e\tau_h$	$\mu\tau_h + e\tau_h$
$\mathcal{L} = 10\,\text{fb}^{-1}$	only DY	DY, $t\bar{t}$, W+jets	DY, $t\bar{t}$, W+jets	DY, $t\bar{t}$, W+jets
NTrees	1500	1300	1600	1600
Shrinkage	0.07	0.17	0.13	0.13
MinNodeSize	4.0%	3.5%	3.5%	3.5%
MaxDepth	3	3	3	3
nCuts	20	40	40	40
UseBaggedSample	✓	✓	✓	✓
BaggedSampleFraction	0.5	0.4	0.6	0.6
BoostType	GradBoost	GradBoost	GradBoost	GradBoost
$m_{\tau\tau}^{SVFit}$	✓	✓	✓	✓
$\Delta R_{lep,\tau}$	✓	✓		
$\Delta\eta_{jj}$		✓	✓	✓
m_{jj}	✓		✓	
$\eta_{j_1} \cdot \eta_{j_2}$				
p_T^{tot}	✓	✓	✓	✓
E_T^{miss} centrality	✓	✓		✓
m_T		✓	✓	✓
η_{lep} centrality	✓	✓	✓	✓
p_T^{sum}	✓	✓	✓	✓
sphericity				
m_{vis}	✓		✓	
AMS	3.7	2.6	2.0	2.9

8.8 Comparison between different BDT parameter sets and input variables for 13 TeV

All BDT parameters and input variable sets of the different analyses performed in this chapter are summarized in Tab. 8.9.

Out of the 12 considered input variables, three are included in each analysis for 13 TeV, namely $m_{\tau\tau}^{SVFit}$, p_T^{sum} and η_{lep} centrality. For the searches including the three background processes DY, $t\bar{t}$ and W+jets, two more variables are present in each BDT analysis, $\Delta\eta_{jj}$ and m_T. Two variables were never included in the optimal variable set, namely η_{lep} centrality and sphericity.

The most significant difference between the optimal BDT parameter sets for the analyses for 13 TeV are given in the values of the Shrinkage parameter. When considering only the DY background process, the parameter is chosen relatively small (0.07), whereas for the three backgrounds DY, $t\bar{t}$ and W+jets it increases. For the $\mu\tau_h$ channel it is set to 0.17, for $e\tau_h$ it is 0.13. The other parameters do not vary significantly. Gradiant boosting and bagging, the use of a random subsample for training, are for all performed analyses the best options in terms of performance. BaggedSampleFraction stays in a range from 0.4 to 0.6 and the values of NTrees are between 1300 to 1600. MaxDepth always has the value 3 and MinNodeSize is set to approximately 3.5%.

9 CMS statistic framework

For the analyses discussed in Chap. 7 and 8 the AMS is used for the quantification of the performance and the comparison between different analysis methods and BDTs. In sec. 5 a second method of quantifying the performance was introduced, the Gaussian significance derived from the corresponding p-value (see sec. 5.3 for detailed explanation).

In the present chapter one possibility of calculating the Gaussian significance of the event classification process is described by the use of the two CMSSW frameworks "CombinedLimit" and "CombineHarvester". The CombinedLimit framework provides a broad range of statistical evaluation methods, which are described in Ref. [39]. It can be used for setting limits on signal processes and calculating the p-value after event classification, which is used in the present case. The framework requires two input files, a datacard of the signal extraction problem and a ROOT file containing event distributions for all processes. The CombineHarvester is an auxiliary tool to create the datacard, which is then used by the CombinedLimit framework.

The CombinedLimit framework further provides the opportunity of including various systematic uncertainties. In the datacard (see sec. 9.2) all signal and background processes included in the analysis are declared as well as all occurring systematic uncertainties that are discussed in sec. 9.1. Systematic uncertainties include normalization uncertainties and shape uncertainties, which require more specific information on the processes. This information is provided in ROOT data files by additional histograms. In sec. 9.1 and 9.3 the production of the ROOT input file is explained in detail.

In the end of this chapter the significance is calculated by CombindLimit for the analyses for 13 TeV including the Drell-Yan, $t\bar{t}$ and W+jets background. Firstly the statistical significance for the combined sample set including both the $\mu\tau_h$ and $e\tau_h$ final states is estimated, followed by the evaluation of the signal extraction process with the two final states being considered as separate channels.

9.1 Systematic uncertainties

The BDT score distribution that is used for the quantification of the BDT performance is strongly affected by imprecisely known quantities. These have several origins. They are either theory related or occur during the experimental process. All of them are combined in the term systematic uncertainties. One has to differentiate between systematic uncertainties that alter the total number of detected events and therefore change the normalization and those that affect the shape of the event distribution, or both.

9.1.1 Normalization uncertainties

In the calculation of the significance several normalization uncertainties are included. They can be caused by uncertainties of the cross section of the processes, the detection process or the beam luminosity in the experiment. The uncertainty in the luminosity is 3% for all processes. Furthermore, each process has cross section uncertainties. The uncertainty for the Drell-Yan $Z \rightarrow \tau\tau$ process is 10%, in $t\bar{t}$ it is 25% and in the signal process $H \rightarrow \tau\tau$ it is given by 20%. In the W+jets background the uncertainty is determined to be 50% [20]. The uncertainty for W+jets also affects the shape and is implemented as shape uncertainty, as explained in the following section.

9.1.2 Shape uncertainties

In addition to normalization uncertainties of the expected number of events, also other uncertainties exist that affect specific kinematic observables like the energy of the particles. These uncertainties are caused by the finite resolution of the detectors or the misidentification of particles and jets. Therefore the total number of expected events may not be affected, but the BDT algorithms gives a different result as the distribution of the input variables of data may differ from the training sample.

In order to provide information on the altered shape, several histograms showing different BDT score distributions need to be provided to the CombindLimit framework. The framework gets all information through the ROOT input file. First of all, the 'original' distribution is included in the ROOT file, which is used for the significance calculations. Then, for each shape uncertainty two more histograms (DOWN and UP) are added to the file, one showing the distribution assuming down-scaled values, the other for up-scaled values.

Two types of shape uncertainties will be discussed. The first ones are caused by identification uncertainties and therefore also lead to a change in the number of expected events. The others alter the kinematic quantities of the events by uncertainties in the energy scale of the particles. For the calculation of the significance performed in this thesis the included shape uncertainties are explained in this chapter.

For gathering information on the energy scale uncertainties, the analysis is done by manipulating the energy of the concerned particle. These manipulations are performed before the final preselection (see Tab. 6.1) is applied on the MC events.

Identification efficiency of real and fake τ_h

The identification and trigger efficiency of real τ_h leads to a total rate uncertainty of 8%, which is relevant for the signal process, the Drell-Yan and the $t\bar{t}$ process. The W+jets background consists exclusively of electrons, muons and jets, which are misidentified as τ_h. The identification of the fake τ_h leads to an estimated rate uncertainty of 50%. [20]

To derive the histograms providing the information on the uncertainty in the identification of real τ_h, randomly 8% of the events of the affected processes (signal, DY, $t\bar{t}$) are

removed form the initial MC samples. The BDT is applied and the BDT score distributions of the processes are used as the DOWN histograms for the particular uncertainty. As there are not more events available than the ones included in the initial MC samples (all of them already used for the analysis), it is not possible to add an amount of 8% additional events. Therefore the UP histograms are created by adding the differences between the original and DOWN histograms to the original histograms, which gives the UP histograms. The same procedure is applied for the identification uncertainty of the fake τ_h (50%) of the W+jets background.

Energy scale uncertainty of τ_h

This uncertainty causes a change in the distribution for several input variables and therefore also a different shape of the BDT score distribution. In particular the uncertainty in the energy scale of each τ_h is given by 3% [20]. In order to extract the additional histograms needed, the energy of the τ_h in each event is rescaled, once with a factor of 1.03 to obtain the UP and once with 0.97 for the DOWN histograms.

This change of the energy affects also other quantities of the event which are relevant for the BDT analysis.

For the τ_h energy scale the following input variables are affected: $m_{\tau\tau}^{SVFit}$, p_T^{sum}, p_T^{tot} and m_{vis}. As the mass estimation of the $\tau\tau$-system is calculated by the highly complex SVFit algorithm, the dependency of $m_{\tau\tau}^{SVFit}$ on E_{τ_h} is approximated by a rescaling factor of 1.015 for UP and 0.975 for DOWN. The p_T^{sum} corresponds to the scalar sum of all transverse momenta and therefore contains the p_T of the lepton (μ or e), the hadronically decaying tau τ_h, the missing transverse energy E_T^{miss} and the two leading jets.

As the original MC samples only include events with $p_T(\tau_h) > 20$ already before the preselection, the preselection cut on $p_T(\tau_h)$ can not be loosened. Therefore only the down-scaling of the τ_h energy makes sense, as for the up-scaling also events that are not available in the MC samples would survive the preselection. The derived BDT score distribution would be biased in the way that events with low $p_T(\tau_h)$ were not included in the analysis.

After the rescaling the original preselection is applied. The derived MC sample is then used for firstly generating the BDT score distribution of the DOWN histograms. By adding the differences between the original and the DOWN histograms to the original histograms, the UP histograms are produced.

Energy scale uncertainty of E_T^{miss}

The reconstruction of E_T^{miss} shows an uncertainty of 5% [20], which leads to the rescaling factors of 1.05 of E_T^{miss} for the UP and 0.95 for the DOWN histograms. The $m_{\tau\tau}^{SVFit}$ is modified by the factor of 1.025 for UP and 0.975 for DOWN. The variables p_T^{sum} and p_T^{tot} are calculated as shown for the τ_h energy scale uncertainty.

In contrast to the τ_h energy scale uncertainty it is possible to extract both the UP and DOWN histograms by applying the BDT on the modified MC samples.

Energy scale uncertainty of jets

The jet energy scale uncertainty concerns all jets included in the MC event and is given by 3% [20]. By modifying the energy scale also the transverse momenta $p_T(jet)$ of the jets change. This affects the p_T^{sum} and p_T^{tot} as they depend on the p_T of the two leading jets. Also the m_{jj} needs to be recalculated.

For the τ_h and E_T^{miss} energy scale uncertainty only the cut $p_T > 30\,GeV$ on the p_T of the respective object (τ_h or E_T^{miss}) is based on the transverse momentum. For the jets also another preselection criteria, namely the central jet veto CJV, introduced in sec. 6.2, depends on $p_T(jet)$.

During original preselection the selection cut $p_T(jet)$ is applied on all jets removing those with a p_T too small, but keeping the event in each case as long as at least to jets survive. Then a central jet veto is applied, which removes the whole event from the analysis if at least one jet has an η with $\eta_1 < \eta < \eta_2$ or $\eta_2 < \eta < \eta_1$. If the preselection cut is now lowered, more jets are kept in the event, which increases the probability of the event having jets with $\eta_1 < \eta < \eta_2$ or $\eta_2 < \eta < \eta_1$. Therefore more events are removed from the sample due to the CJV. The final MC samples contain fewer events caused by a looser p_T selection cut.

9.2 Datacard

The datacard is produced by the CombineHarvester framework. To generate the datacard all event categories, processes and uncertainties must be declared. Several parameters are provided by the tool that need to be filled accordingly.

Firstly, the categories that declare the production process and the final state need to be set. In the particular case either one or two are created named according to the final state that is analysed, namely `muTau_vbf` and/or `eTau_vbf`, which also get assigned a `bin_id` of 1 and/or 2. More information is provided in the datacard by additional parameters that are defined, namely `analysis="htt"`, `era="13TeV"`, `channel="mt"` or `"et"` and the `mass="125"`, which corresponds to the Higgs mass and is only set for the signal process. Then the processes are defined and split into signal and background processes: `sig_procs` only contains the signal `"sgn"`, whereas `bkg_procs` contains `"DY"`, `"ttbar"` and `"Wjets"`.

For the systematic uncertainties two different types must be distinguished. The normalization uncertainties are declared by using the distribution type `"lnN"`, which corresponds to a log-normal distribution and is recommended for multiplicative corrections like normalization. All processes that are affected by a particular normalization uncertainty get assigned the size of the uncertainty. The value is defined by $1 + \Delta x/x$, with $\Delta x/x$ corresponding to the relative uncertainty. Four different norm uncertainties are

included, namely `lumi` with the factor 1.03, which affects all processes, and the uncertainties for each specific process, namely `sgn125_norm` with a factor of 1.2, `DY_norm` with 1.1 and `ttbar_norm` with 1.25. All of them are introduced in sec. 9.1.1.

The second type is given by the shape uncertainties. In this case the distribution type must be set to `"shape"`, which tells the CombinedLimit framework that there are additional histograms available in the ROOT input file, that give information on the impact of various uncertainties on the event distribution. These shape uncertainties always affect the distribution of all processes. The uncertainties included are given by `"IDfakeTau"`, `"IDrealTau"`, `"tau_energy"`, `"met_energy"` and `"jet_energy"`, which are discussed in section 9.1.2.

Additionally, the datacard displays the number of observed events for each process, which corresponds to the weighted number of MC events. These numbers are automatically extracted from the ROOT input file as well as the number of bins and processes (imax and jmax). The number of nuisance parameters, denoted as kmax, is not stated in the datacard, but is directly extracted from the ROOT file by the CombinedLimit tool.

For each analysis carried out a separate datacard is generated that matches the present event classification problem. Two examples are provided below, the first one shows the datacard for the $\mu\tau_h$ analysis, the second one the combination of the $\mu\tau_h$ and the $e\tau_h$ final states considered as separate channels.

htt_mt_1_13TeV_125.txt

```
# Datacard produced by CombineHarvester with git status: CombineHarvester-v15.5-13-g9c77453
imax    1 number of bins
jmax    3 number of processes minus 1
kmax    * number of nuisance parameters
-------------------------------------------------------------------------------------
shapes * htt_mt_1_13TeV htt_mt.input.root htt_mt_1_13TeV/$PROCESS htt_mt_1_13TeV/$PROCESS_$SYSTEMATIC
shapes sgn htt_mt_1_13TeV htt_mt.input.root htt_mt_1_13TeV/sgn$MASS htt_mt_1_13TeV/sgn$MASS_$SYSTEMATIC
-------------------------------------------------------------------------------------
bin           htt_mt_1_13TeV
observation   4785.4404
-------------------------------------------------------------------------------------
bin                       htt_mt_1_13TeV   htt_mt_1_13TeV   htt_mt_1_13TeV   htt_mt_1_13TeV
process                   sgn              DY               ttbar            Wjets
process                   0                1                2                3
rate                      21.51            963.8            817              2862
-------------------------------------------------------------------------------------

DY_norm        lnN      -                1.1              -                -
IDfakeTau_     shape    -                -                -                1
IDrealTau_     shape    1                1                1                -
jet_energy_    shape    1                1                1                1
lumi           lnN      1.03             1.03             1.03             1.03
met_energy_    shape    1                1                1                1
sgn125_norm    lnN      1.2              -                -                -
tau_energy_    shape    1                1                1                1
ttbar_norm     lnN      -                -                1.25             -
```

```
                                    combined_125.txt

# Datacard produced by CombineHarvester with git status: CombineHarvester-v15.5-13-g9c77453
imax    2 number of bins
jmax    3 number of processes minus 1
kmax    * number of nuisance parameters
----------------------------------------------------------------------------------
shapes * htt_et_2_13TeV htt.input.root htt_et_2_13TeV/$PROCESS htt_et_2_13TeV/$PROCESS_$SYSTEMATIC
shapes sgn htt_et_2_13TeV htt.input.root htt_et_2_13TeV/sgn$MASS htt_et_2_13TeV/sgn$MASS_$SYSTEMATIC

shapes * htt_mt_1_13TeV htt.input.root htt_mt_1_13TeV/$PROCESS htt_mt_1_13TeV/$PROCESS_$SYSTEMATIC
shapes sgn htt_mt_1_13TeV htt.input.root htt_mt_1_13TeV/sgn$MASS htt_mt_1_13TeV/sgn$MASS_$SYSTEMATIC
----------------------------------------------------------------------------------
bin         htt_et_2_13TeV   htt_mt_1_13TeV
observation 2624.8958        4785.4404
----------------------------------------------------------------------------------

bin                   htt_et_2_13TeV  htt_mt_1_13TeV  htt_et_2_13TeV  htt_et_2_13TeV  htt_et_2_13TeV  htt_mt_1_13TeV  htt_mt_1_13TeV  htt_mt_1_13TeV
process               sgn             sgn             DY              ttbar           Wjets           DY              ttbar           Wjets
process               0               0               1               2               3               1               2               3
rate                  12.47           21.51           243             461.8           1798            963.8           817             2862
----------------------------------------------------------------------------------
DY_norm       lnN     -               -               1.1             -               -               1.1             -               -
IDfakeTau_    shape   -               -               -               -               1               -               -               1

IDrealTau_    shape   1               1               1               1               -               1               1               -

jet_energy_   shape   1               1               1               1               1               1               1               1
lumi          lnN     1.03            1.03            1.03            1.03            1.03            1.03            1.03            1.03
met_energy_   shape   1               1               1               1               1               1               1               1
sgn125_norm   lnN     1.2             1.2             -               -               -               -               -               -
tau_energy_   shape   1               1               1               1               1               1               1               1
ttbar_norm    lnN     -               -               -               1.25            -               -               1.25            -
```

9.3 ROOT input file

The ROOT input file provides the histograms for the calculation of the expected signal significance. In the present case the histograms show the BDT score distribution of the MC events. The ROOT file is named according to the analysis, e.g. `htt.input.root` (including $\mu\tau_h$ and $e\tau_h$) or `htt_mt.input.root` if only one channel is considered.

All histograms of one category are arranged in a folder named accordingly to the category (`htt_mt_1_13TeV` or `htt_et_2_13TeV`). In each folder several histograms are stored. The most important one is called `"data_obs"`, which normally provides the distribution of the actual data taken from the experiment. So far no data is analysed, but the analysis method is tested using MC samples. Therefore the `"data_obs"` histogram includes the MC events of all processes. Furthermore the distribution of each process is shown in a separate histogram, namely `"sgn125"`, `"DY"`, `"ttbar"` and `"Wjets"`. The names must correspond to the names of the processes declared in the datacard. For the shape uncertainties additional histograms are added for each process. For instance the signal process is affected by the uncertainties in the real τ_h identification, the jet energy scale, the E_T^{miss} energy scale and the τ_h energy scale. The datacard shows which shape uncertainty affects which process. For each uncertainty, one UP and one DOWN shape histogram is added to the ROOT file. This sums up to one histogram showing the "observed data", 1+4x2 histograms for the signal process, 1+4x2 for DY, 1+4x2 for $t\bar{t}$ and 1+4x2 for the W+jets background. In total this leads to 37 histograms in each category folder. In the ROOT input file all histograms show the same binning, namely 21 bins in the BDT score range [-1,1].

9.4 Results of the 13 TeV analysis

As discussed above, the Gaussian significance is calculated by the CMS CombinedLimit framework. The significance is firstly calculated for the single $\mu\tau_h$ channel and the single $e\tau_h$ channel leading to the values 2.3 and 2.0.

The combination of these two channels is performed in two different ways, similar to sec. 8.7. In the first case, both channels are considered separately and two categories are defined in the datacard. This leads to a significance of 3.4, denoted as $\mu\tau_h\_e\tau_h$.

In the second case, the BDT, which is trained on events from both channels, is used to generate the histograms for the ROOT input file. Here, the two final states are considered as one channel, which means that also in the datacard only one channel (one category) is defined. The evaluation of this analysis ($\mu\tau_h + e\tau_h$) gives a significance of 4.0. This value is significantly higher than the significance of $\mu\tau_h\_e\tau_h$ and seems to be overestimated compared to the other values. This overestimation may be a consequence of the fact that statistical uncertainties are not included in CombinedLimit calculations.

In Tab. 9.1 the AMS and the significance calculated by the CombinedLimit tool for all analyses considering all three backgrounds are summarized. It must be borne in

mind, that these quantifications of the performance can not exactly be compared, as different uncertainties are included in each evaluation method. The calculation of the AMS considers the statistical uncertainties of each bin, whereas these uncertainties are not included in the significance derived by CombinedLimit. In case of the systematic uncertainties, only the normalization uncertainties are taken under consideration in the AMS. The significance on the other hand also includes shape uncertainties.

Overall, the AMS and the CombinedLimit significance agree reasonably. For both quantities the $\mu\tau_h$ channel shows better results than the $e\tau_h$ channel and the combined analyses give higher values than the single analyses. Nevertheless, it is essential to include statistical uncertainties in the evaluation of the significance calculations of the CombinedLimit tool for further investigations.

Tab. 9.1: Summary of the AMS and the significances calculated by CombinedLimit for different analyses. $\mu\tau_h\_e\tau_h$ corresponds to two separate BDT classifiers, whereas $\mu\tau_h + e\tau_h$ corresponds to one BDT trained on events from both channels.

Channels	AMS	Gaussian significance
$\mu\tau_h$	2.6	2.3
$e\tau_h$	2.0	2.0
$\mu\tau_h\_e\tau_h$	2.9	3.4
$\mu\tau_h + e\tau_h$	2.9	4.0

10 Conclusions and outlook

This thesis presents a study on the signal extraction performance of Boosted Decision Trees for the $H \to \tau^+\tau^-$ channel. The search includes events of a Higgs boson produced by vector-boson fusion decaying to a pair of τ making use of the final states $\mu\tau_h$ and $e\tau_h$. The background treatment is performed by implementation of the irreducible Drell-Yan background process ($Z \to \tau^+\tau^-$) and the reducible $t\bar{t}$ and W−jets processes.

The aim of the study is the optimization of a BDT method for the signal extraction for 8 and 13 TeV, with a focus on the 13 TeV analysis. Both analyses are performed by utilizing simulated Monte Carlo samples. The MC samples used for the 8 TeV analysis are normalized to a luminosity of $\mathcal{L} = 19.7\,\mathrm{fb}^{-1}$ and are provided for the signal process and the Drell-Yan background both for the final state $\mu\tau_h$. For 8 TeV, the optimized BDT algorithm performs significantly better (AMS 3.5) than the cut-based analysis (AMS=2.1). The absolute numbers are only indicative as many approximations are made, like the simplified statistics and systematics treatment and the consideration of only DY. However, the BDT analysis performs significantly better than the cut-based.

For 13 TeV, the BDT is firstly trained and optimized on the MC samples for the signal and DY processes for the final state $\mu\tau_h$ normalized to $\mathcal{L} = 10\,\mathrm{fb}^{-1}$. The choice of the starting point of the optimization procedure is made considering the insights gained in the 8 TeV analysis. The optimized BDT gives an AMS of 3.7, compared to 1.9 from the cut-based analysis.

For the following analyses results, two types of quantifying the performance are conducted. The value stated first is the approximate median significance, the second value stated in brackets denotes the significance calculated by the CMS CombinedLimit framework. For a more complete background treatment, the $t\bar{t}$ and W+jets processes are implemented. Under consideration of these three backgrounds the optimization of the BDT algorithm is performed again. The performance of the optimized BDT is given by a significance of 2.6 (2.3), whereas an AMS of 1.05 is obtained for the cut-based analysis. For the analysis of the $e\tau_h$ the BDT is reoptimized leading to a significance of 2.0 (2.0). As the signal yield in this channel is lower than in $\mu\tau_h$, a lower AMS in $e\tau_h$ was expected.

To make use of the signal extraction performance of both channels, two different approaches for combining the two channels are studied. In the first approach ($\mu\tau_h\_e\tau_h$) the event classification is performed separately for each channel. The AMS is calculated from the sum of the events of both channels. In the CombinedLimit framework, two separate channels are declared. The significance is calculated to 2.9 (3.4). For the second approach the BDT algorithm is reoptimized and trained only once using the MC samples for both the $\mu\tau_h$ and $e\tau_h$ channel. This leads to a significance of 2.9 (4.0). The AMS does

not include shape uncertainties, which are considered in CombinedLimit. Due to these differences, the values of the differently calculated significances can not be compared directly.

The study of the $\mu\tau_h$ channel for 13 TeV only including the DY background is performed with different analyses methods. Besides BDT, the performance of a cut-based approach and two multivariate analyses that are not part of this thesis, namely NeuroBayes [38] and deep neural networks [17], is investigated. A comparison shows, that the BDT analysis performs significantly better than the other approaches. An additional advantage of a BDT is its good "out-of-the-box" performance and it requires relatively low effort for optimization compared to deep neural networks.

This gives evidence that a BDT analysis has high potential to enhance the performance of the signal extraction when applied on the 13 TeV data. CMS is currently in the data taking process. In 2016, enough data will be available to perform the $H \rightarrow \tau^+\tau^-$ analysis utilizing the optimized BDT algorithms extracted in the course of this thesis. Before this analysis can be done, several tasks need to be accomplished. Firstly, a full implementation of the CMS Statistic framework (CombinedLimit) is required, including the statistical uncertainties. Secondly, the background treatment must be completed. A validation on the background modelling must be performed as well as the background estimation techniques based on data control regions will be updated.

11 Acknowledgement

At this point I want to mention some people that played a significant role during the time of my thesis and my studies in general:

I want to thank Manfred Jeitler, my university supervisor, for a pleasant cooperation and stimulating discussions during my time at CERN.

Many thanks to Martin Flechl for being a great supervisor, introducing me to this interesting topic, supporting and promoting me all the time during my work and giving me interesting tasks and problems to solve. I learned a lot during the last months.

Special thanks to Wolfgang Adam, the team leader of the CMS analysis group at HEPHY Vienna, and Martin Flechl for making it possible for me to participate in the CMS week at CERN in October this year. It was an extraordinary experience to get an insight in the incredible work that is done there every day and the problems that need to be worked out in this large and international collaboration.

Thanks to Johannes Brandstetter for taking his time to answer many questions, for struggling with and solving out problems together and last but not least for a great week in Geneva and its surroundings.

I also want to thank all my colleagues at HEPHY Vienna for a great working environment and many interesting and amusing discussions during lunch time and coffee breaks or in the office. I really enjoyed the time at HEPHY.

I am very grateful to my parents who supported me mentally and financially, not only throughout this thesis but during my whole studies. I want to thank my father, Herbert, for many lively discussions on physics related problems, and my mother, Gerlinde, and my sister, Carina, for having the patience to bear these conversations.

Last but not least I want to thank all my friends, the new and old ones, for making the years of study an incredible and unforgetable time of my life.

Appendices

A Acronyms

ALICE	A Large Ion Collider Experiment	**HE**	Hadronic Endcap Calorimeter
AMS	Approximate median significance	**HF**	Hadronic Forward Calorimeter
APD	Avalanche Photodiode	**HLT**	High-Level Trigger
ATLAS	A Toroidal LHC ApparatuS	**HO**	Hadronic Outer Calorimeter
BDT	Boosted Decision Tree	**L1**	Level-1 Trigger
BR	Branching Ratio	**LEAR**	Low Energy Antiproton Ring
CJV	Central Jet Veto	**LEP**	Large Electron-Positron Collider
CMS	Compact Muon Solenoid	**LHC**	Large Hadron Collider
CP	Charge conjugation, Parity transformation	**LHCb**	Large Hadron Collider beauty
CPT	Charge conjugation, Parity transformation, Time reversal	**LINAC**	Linear Accelerator
		MC	Monte Carlo
CSC	Cathode Strip Chambers	**ML**	Maximum-Likelihood
DAQ	Data Acquisition	**MVA**	Mulivariate Analysis
DT	Drift Tube	**pdf**	Probability Density Function
DY	Drell-Yan	**PS**	Proton Synchrotron
ECAL	Electromagnetic Calorimeter	**PSB**	Proton Synchrotron Booster
FoM	Figure of Merit	**QCD**	Quantum chromodynamics
ggF	gluon-gluon Fusion	**QED**	Quantum electrodynamics
GUT	Grand Unified Theory	**RPC**	Resistive Plate Chambers
HCAL	Hadronic Calorimeter	**SM**	Standard Model of particle physics
HB	Hadronic Barrel Calorimeter	**SPS**	Super Proton Synchrotron
		SU	Special Unitary Group

SUSY	Supersymmetry	**VBF**	Vector-Boson Fusion
TMVA	Toolkit for Multivariate Analysis	**XS**	Cross Section
U	Unitary Group		

Bibliography

[1] W.N. Cottingham and D.A. Greenwood. *An Introduction to the Standard Model of Particle Physics*. Cambridge University Press, 2007.

[2] P.W. Higgs. Broken symmetries, massless particles and gauge fields. *Physics Letters*, 12(2):132 − 133, 1964.

[3] P. W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964.

[4] P. W. Higgs. Spontaneous symmetry breakdown without massless bosons. *Phys. Rev.*, 145:1156–1163, May 1966.

[5] UA1 Collaboration. Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at s**(1/2) = 540-GeV. *Phys. Lett.*, B122:103–116, 1983.

[6] UA1 Collaboration. Studies of Intermediate Vector Boson Production and Decay in UA1 at the CERN Proton - Antiproton Collider. *Z. Phys.*, C44:15–61, 1989.

[7] CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett.*, B716:30–61, 2012.

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

[9] ATLAS & CMS Collaboration. Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments. *Phys. Rev. Lett.*, 114:191803, 2015.

[10] H. YUKAWA. On the interaction of elementary particles. I. *Proceedings of the Physico-Mathematical Society of Japan. 3rd Series*, 17:48–57, 1935.

[11] M. Carena, C. Grojean, M. Kado, and V. Sharma. *Status of Higgs boson physics*. Particle Data Group, 2013.

[12] MissMJ. Standard model of elementary particles. https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg, June 2006. Accessed: 2015-11-27.

[13] LHCb Collaboration. Observation of $j/\psi p$ resonances consistent with pentaquark states in $\Lambda_b^0 \to j/\psi K^- p$ decays. *Phys. Rev. Lett.*, 115:072001, Aug 2015.

[14] CMS Collaboration. The Discovery of the Higgs Boson with the CMS Detector and its Implications for Supersymmetry and Cosmology. In *Time and Matter 2013 (TAM2013) Venice, Italy*, 2013.

[15] Y. Nambu. Quasi-particles and gauge invariance in the theory of superconductivity. *Phys. Rev.*, 117:648–663, Feb 1960.

[16] J. Goldstone. Field theories with « superconductor » solutions. *Il Nuovo Cimento (1955-1965)*, 19(1):154–164, 1961.

[17] M. Spanring. *Study of potential improvements of the CMS $H \to \tau^+\tau^-$ analysis using artificial neural networks with multiple layers*. Vienna University of Technology, 2015.

[18] A. Denner, S. Heinemeyer, I. Puljak, D. Rebuzzi, and M. Spira. Standard Model Higgs-Boson Branching Ratios with Uncertainties. *Eur. Phys. J.*, C71:1753, 2011.

[19] ATLAS Collaboration. Evidence for the Higgs-boson Yukawa coupling to tau leptons with the ATLAS detector. *JHEP*, 04:117, 2015.

[20] CMS Collaboration. Evidence for the 125 GeV higgs boson decaying to a pair of τ leptons. *Journal of High Energy Physics*, 2014(5), 2014.

[21] CMS Collaboration. Reconstruction and identification of τ lepton decays to hadrons and ν_τ at CMS. *ArXiv e-prints*, 2015.

[22] CERN. The history of cern. http://timeline.web.cern.ch/timelines/The-history-of-CERN. Accessed: 2015-12-01.

[23] L. Evans and P. Bryant. LHC Machine. *JINST*, 3:S08001, 2008.

[24] CMS Collaboration. *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. Technical Design Report CMS. CERN, Geneva, 2006.

[25] CERN. Accelerators. http://home.cern/about/accelerators-topics. Accessed: 2015-12-01.

[26] J. Timmer. Following protons on a trip to (and through) the LHC. http://arstechnica.com/science/2010/08/following-protons-on-a-trip-to-and-through-the-lhc/, Aug 2010. Accessed: 2015-11-30.

[27] ALICE Collaboration. The ALICE experiment at the CERN LHC. *JINST*, 3:S08002, 2008.

[28] ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008.

[29] CMS Collaboration. The CMS experiment at the CERN LHC. *JINST*, 3:S08004, 2008.

[30] LHCb Collaboration. The LHCb Detector at the LHC. *JINST*, 3:S08005, 2008.

[31] T. Sakuma and T. McCauley. Detector and Event Visualization with SketchUp at the CMS Experiment. *J. Phys. Conf. Ser.*, 513:022032, 2014.

[32] CMS Collaboration. CMS detector overview. http://cmsinfo.web.cern.ch/ cmsinfo/Detector/FullDetector/index.html. Accessed: 2015-11-30.

[33] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss. TMVA - Toolkit for Multivariate Data Analysis. *ArXiv Physics e-prints*, March 2007.

[34] R. Brun and F. Rademakers. ROOT: An object oriented data analysis framework. *Nucl. Instrum. Meth.*, A389:81–86, 1997.

[35] J. H. Friedman. Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 38(4):367–378, 2002.

[36] G. Cowan and E. Gross. Discovery significance with statistical uncertainty in the background estimate. *URL http://www. pp. rhul. ac. uk/~ cowan/stat/notes/SigCalcNote. pdf*, 2008.

[37] G Cowan, K. Cranmer, E. Gross, and O. Vitells. Asymptotic formulae for likelihood-based tests of new physics. *The European Physical Journal C*, 71(2), 2011.

[38] S. Kloibhofer. Potential of the analysis of higgs boson decays to two tau leptons with neurobayes with the CMS experiment, 2015.

[39] ATLAS Collaboration. Procedure for the LHC Higgs boson search combination in summer 2011. *ATL-PHYS-PUB-2011-11*, 2011.