Search for Diboson Production in 4τ Final States with the ATLAS Detector

Dominik Weiß

Masterarbeit in Physik
angefertigt im Physikalischen Institut

vorgelegt der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität
Bonn

Oktober 2021

I hereby declare that this thesis was formulated by myself and that no sources or tools other than those cited were used.

Bonn, ...

Date Signature

1. Gutachter: Prof. Dr. Klaus Desch
2. Gutachter: Prof. Dr. Jochen Dingfelder

Acknowledgments

I would like to thank Prof. Dr. Klaus Desch for giving me the possibility of writing my Master thesis in his research group. Furthermore, I want to thank Dr. Philip Bechtle and Dr. Christian Grefe for all their ideas and ongoing help in a lot of interesting discussions. The same applies to Henrik Junkerkalefeld who was a great support for my questions in the first months as well.
In addition, especially Lena Herrmann, Jan-Eric Heinrichs, Marco Menen and Tobias Klingl but of course also the remaining people of the research group deserve a big thank-you for all of their help and the great working atmosphere including office sport sessions, occasional chess games, lunch breaks often filled with funny discussions and in general the productive and very enjoyable time in the office.
Finally, I would like to thank my family, further friends and flatmates for their ongoing support during the last year and the whole study time.

Contents

1 Introduction **1**

2 The Standard Model **3**
2.1 Particles and Fundamental Forces . . . 3
2.2 Electroweak Interaction . . . 5
2.3 Quantum Chromodynamics . . . 7
2.4 The Z Boson . . . 8
2.4.1 Properties and Decay . . . 8
2.4.2 ZZ Diboson Production . . . 9
2.5 Higgs Mechanism . . . 11
2.5.1 Electroweak Symmetry Breaking and the Higgs Field . . . 11
2.5.2 The Higgs Boson . . . 13
2.6 The τ-Lepton . . . 13

3 Experimental Setup **15**
3.1 The Large Hadron Collider . . . 15
3.2 The ATLAS Detector . . . 16
3.2.1 ATLAS Coordinate System . . . 17
3.2.2 Triggers . . . 18
3.2.3 Tracking & Vertexing in the Inner Detector . . . 19
3.2.4 Calorimetry . . . 20
3.2.5 Muon System . . . 21
3.3 Object Reconstruction at ATLAS . . . 21
3.3.1 Jets . . . 21
3.3.2 Leptons & Photons . . . 22
3.3.3 Missing Transverse Energy . . . 23
3.4 Event Generation & Monte Carlo Simulation . . . 23

4 TMVA & Boosted Decision Trees **27**
4.1 Training using the TMVA Factory . . . 27
4.2 Evaluation using the TMVA Reader . . . 28
4.3 Decision Trees . . . 29
4.3.1 Basics . . . 29
4.3.2 Boosting . . . 30

5 Multi-Tau Event Identification **33**
5.1 Tau Lepton Identification in ATLAS . . . 33
5.2 Multiplied Probability Method . . . 37

6 Data Samples & Background Estimation **41**
6.1 Observable Definitions . . . 41
6.2 Signal Processes . . . 42
6.3 Monte Carlo Background Processes . . . 43
6.4 Preselection . . . 44
6.5 Data Driven Estimation of Fake Taus from QCD Multijet Processes . . . 46
6.6 Advanced Data Driven Fake Tau Background Estimation . . . 47

7 Event Selection in the $ZZ \rightarrow 4\tau$ Channel **51**
7.1 Cut-Based Approach . . . 51
7.2 Multivariate Approach using Boosted Decision Trees . . . 54
7.2.1 Training Samples and Variables . . . 54
7.2.2 Training Results . . . 55
7.2.3 Monte Carlo BDT Performance . . . 60
7.2.4 QCD Multijet BDT Performance . . . 62
7.3 Combination of Tau ID and BDT Classifier Output . . . 65
7.3.1 Control Regions . . . 68
7.3.2 Signal Regions . . . 69
7.3.3 Validation Region . . . 70

8 Measuring the $ZZ \rightarrow 4\tau$ Cross Section **73**
8.1 Profile Likelihood Fit Procedure . . . 73
8.2 Systematic Uncertainties . . . 74
8.2.1 Tau ID Uncertainties . . . 74
8.2.2 Further Uncertainties . . . 75
8.2.3 Pruning . . . 76
8.3 Fit Results . . . 77
8.3.1 Asimov Data . . . 77
8.3.2 ATLAS Data . . . 80
8.4 Interpretation in the Standard Model . . . 84

9 Conclusion **87**

Bibliography **89**

A Detailed List of Used Monte Carlo Samples **97**

B Input Variable Distributions for the BDT Training **101**

C Full Ranking Tables of the Input Variables for the BDT Training **109**

D BDT Overtraining Results for Different Hyperparameter Setups **113**

E Pre-Fit Event Yield Tables 119

F Post-Fit Event Yield Tables 123

List of Figures 125

List of Tables 129

CHAPTER 1

Introduction

Humans did always try to understand and describe scientific phenomena by creating models based on their assumptions or experimental observations. They quickly came to the conclusion that science itself is very likely to be infinite, as already stated by the German explorer Alexander von Humboldt [1]:

> "Jedes Naturgesetz, das sich dem Beobachter offenbart, läßt auf ein höheres, noch unerkanntes schließen."

Starting with the discovery of the electron in 1897 [2], the field of atomic, nuclear and particle physics was developed, accompanied by further revolutionary theories like Einstein's special relativity (1905) [3][4], the Bohr model (1913) [5] or the Schrödinger equation (1926) [6] as one of the foundations of quantum mechanics. Nowadays, the Standard Model (SM) of particle physics formulated in the 1960's summarizes the current knowledge in this field of physics (cf. chap. 2). Its predictions have been tested over many decades in experiments using high-energy particle colliders (cf. chap. 3) and most of them were found to be in very good agreement with the observations. In 2012, finally the last particle predicted by the Standard Model was found with the discovery of the Higgs boson at CERN [7]. However, the SM does not give an answer to all questions in particle physics. For example, it does not include a description of the gravitational force or the possible existence of dark matter in the universe and also the origin of the non-zero masses of neutrinos, which are predicted to be massless in the Standard Model, is beyond its scope.
Apart from these distinct hints for physics exceeding the theory of the SM, there are also some processes within it which have not been measured yet. One of them is the $ZZ \rightarrow 4\tau$ decay, where the Z boson is one of the exchange particles of the weak interaction. Albeit it is described by the Standard Model, such a final state with four charged leptons is quite exotic since it can be produced by hardly any other SM process and thus also could be sensitive to possible beyond Standard Model (BSM) physics, e.g. to heavy resonances decaying into two on-shell Z bosons. As the cross section of the $ZZ \rightarrow 4\tau$ process is expected to be very small, a best possible reconstruction and identification of the short-lived τ-leptons emerging from the decay as well as their discrimination from simultaneously produced background objects in the detector is crucial. For that purpose, the features of boosted decision trees (cf. chap. 4), which are a commonly used type of multivariate classifiers in high energy particle physics, are combined with the power of a new identification method for multi-tau events (cf. chap. 5) developed and evaluated on the $hh \rightarrow 4\tau$ process in [8]. Instead of treating each τ-lepton on its own, events of interest are selected based on a multiplied probability in which the

information obtained from the single taus is merged. However, despite its unique final state structure, the $ZZ \rightarrow 4\tau$ channel suffers from a large hadronic background originating from QCD jets which are misidentified as τ-leptons. For the estimation of this background, a data driven technique is adapted from [8] and further extended in chapter 6. Subsequently, an event selection for the $ZZ \rightarrow 4\tau$ channel is designed which exploits the potential of combining the introduced tau identification method with the classification power of boosted decision trees (cf. chap. 7). This selection is then used in chapter 8 to perform a profile likelihood fit in order to measure the $ZZ \rightarrow 4\tau$ cross section compared to the Standard Model prediction, including the effects of systematic uncertainties. Finally, the obtained results are summarized in chapter 9 and an outlook towards further possible studies is given.

CHAPTER 2

The Standard Model

The Standard Model of particle physics describes the elementary particles of our universe and the fundamental forces acting between them. Many of its parameters have been measured with impressively high precision at the end of the last century, e.g. with the Large Electron-Positron (LEP) collider at CERN [9], before the SM was finally completed with the discovery of the Higgs boson in 2012 [7]. Up to now, the Standard Model predictions are in very good agreement with experimental data collected over many decades and represent one of the largest successes of modern physics [10]. Section 2.1 gives a quick overview of the particles and forces in the SM, followed by a short explanation of the underlying concepts of electroweak unification in section 2.2 and quantum chromodynamics in section 2.3. The Z boson and its properties are summarized in section 2.4, before the Higgs mechanism is introduced in section 2.5. Finally, the description is completed with the properties of τ-lepton in section 2.6 which will be of particular interest for the analysis presented in this thesis.

2.1 Particles and Fundamental Forces

In the Standard Model, there are three forces (electromagnetic, weak and strong) acting on particles by the exchange of spin-1 vector bosons, also called force carriers, which are summarized in table 2.1. The additional fourth force, gravity, is not included in the model since its relative strength for elementary particles is many orders of magnitude lower than those of the other forces. The mass of the exchange boson directly affects the range of the corresponding interaction. Photons are massless which leads to an infinite range of the electromagnetic force where only the interaction strength decreases with rising distance between two particles. In contrast to that, the force carriers of the weak interaction have comparably high masses of $m_W = 80.385\,\text{GeV}$ and $m_Z = 91.1876\,\text{GeV}$ (cf. tab. 2.1) resulting in an interaction range of only $\approx 10^{-3}$ fm. Being the exchange particle of the strong interaction, the gluon is massless again. But since it carries color charge itself (unlike the photon which is electrically neutral), it has the possibility to interact with other gluons. The combination of this behavior with the theory of *color confinement*[1] limits the effective range of the strong interaction. Particles participating in this interaction are called hadrons [11].

[1] "Free" particles always need to be color neutral. For distances of more than ≈ 1 fm there is a sufficient amount of energy in a two-particle system to create quark-antiquark pairs from the vacuum which form color neutral hadrons again.

Interaction	Exchange particle	Coupling to	Mass [GeV]
electromagnetic	photon (γ)	electric charge q	0
strong	8 gluons (g)	color charge (red, green, blue)	0
weak	$W^{\pm}$ and Z boson	weak isospin I_W	80.385 / 91.1876

Table 2.1: The exchange particles of the different interactions in the Standard Model and their properties. From [11] chap. 13, masses from [11] chap. 12.1.

Figure 2.1: The Standard Model of particle physics. The three generations of fermions (quarks & leptons) as well as the vector (or gauge) bosons and the Higgs boson including their mass, spin and electric charge are shown. Image from [12].

In addition to these bosons, there are 13 other elementary particles predicted by the Standard Model shown in figure 2.1 which differ in their quantum numbers (e.g. electric charge, flavor, baryon or lepton number). Those are namely the quarks, leptons and the Higgs boson which occurs as excitation of the Higgs field (cf. sec. 2.5). Quarks belong to the group of fermions, i.e. are spin-$\frac{1}{2}$ particles and take part in all SM interactions since they carry electric and color charge as well as an weak isospin unequal to zero. Leptons, being fermions too, only interact weakly and, in case of the charged leptons, also electromagnetically. The Higgs boson couples exclusively to non-massless particles [11].
The fermions of the Standard Model are split in three generations. The first one contains the up and the down quark, which form the proton and the neutron[2], the electron and the electron neutrino. Apart from the last one, these particles represent the elementary building blocks of all visible matter. The higher generation particles differ in their flavor and mass from their first generation partners, but all others quantum numbers are equal. They are unstable and naturally do not occur on earth but can be produced in collisions of high energy particles, e.g. in a particle accelerator. Finally, for each quark

[2] The single constituents of the proton are labeled as partons in the later part of this thesis.

and lepton an antiparticle exists. It has the same mass as the corresponding particle, but opposite quantum numbers [11].

2.2 Electroweak Interaction

Quantum Electrodynamics

Quantum Electrodynamics (QED) is the quantum field theory[3] of the electromagnetic interaction and should be introduced very briefly here. Calculated from time-ordered perturbation theory, it describes fermion interactions of e.g. the type $a + b \longrightarrow c + d$ which take place by the exchange of a particle X. In QED, this mediator X is the photon γ, given by the free photon field [10]

$$A_\mu = \varepsilon_\mu^{(\lambda)} e^{i(\vec{p}\cdot\vec{x} - Et)} \tag{2.1}$$

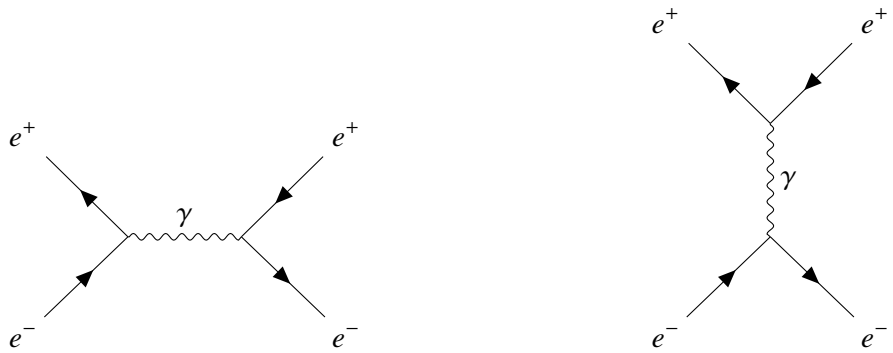
where $\varepsilon^{(\lambda)}$ denotes the four-vector for the polarization state λ. For a real photon, the polarization vector needs to be transverse to the propagation direction, resulting in polarization states of [10]

$$\varepsilon^{(1)} = (0, 1, 0, 0) \quad \text{or} \quad \varepsilon^{(2)} = (0, 0, 1, 0) \tag{2.2}$$

assuming a movement in z-direction. As it can be inferred from above, the fundamental interaction always occurs between two spin-$\frac{1}{2}$ fermions and a single photon. Therefore, QED vertices which connect more than three particles are not allowed and all possible Feynman diagrams are based on this three-particle vertex. In QED processes, each of these vertices contributes a factor of

$$\alpha_{\text{em}} = \frac{e^2}{4\pi} \approx \frac{1}{137} \tag{2.3}$$

to the interaction matrix element where α_{em} is the dimensionless electromagnetic coupling constant [10]. An example for such a Feynman diagram is depicted in figure 2.2, showing electron-positron scattering at tree level (also called Bhabha scattering), i.e. without taking loop effects into account.



(a) s-channel (b) t-channel

Figure 2.2: Feynman diagrams of the Bhabha scattering process, split into s-channel (annihilation) and t-channel (scattering) [14].

[3] For a more detailed introduction of quantum field theory concepts see [10] or [13].

Weak Interaction

In contrast to QED and QCD (cf. sec. 2.3), the weak interaction is mediated by massive spin-1 bosons and is the only Standard Model interaction which does not conserve parity[4]. These properties require the weak interaction to have a different form compared to QED and QCD. Classifying physical quantities and interactions according to table 2.2, the latter interactions fall into the category of vector current interactions with matrix elements being invariant under Lorentz transformations (see [10] for a more detailed explanation). Demanding this invariance for the weak interaction as well, it needs to have a form built from a linear combination of vector (V) and axial vector (A) currents, j_V^μ and j_A^μ, to still allow for the violation of parity [10]:

$$j^\mu \propto g_V j_V^\mu + g_A j_A^\mu \ . \tag{2.4}$$

Here, g_V and g_A denote the vector and axial vector coupling constants, respectively. The relative strength of the parity violating part of the matrix element compared to the part which conserves parity follows as [10]:

$$\frac{g_V g_A}{g_V^2 + g_A^2} \ . \tag{2.5}$$

This allows for the conclusion that parity is conserved for either g_V or g_A being zero, while it is maximally violated if $|g_V| = |g_A|$. Experimental results show that the charged weak interaction mediated by the W boson seems to fulfill this last criterion, leading to a maximally parity violating $V - A$ structure. As a direct consequence, only left-handed chiral[5] particle and right-handed chiral antiparticle states participate in the weak charged-current [10].

Quantity/Interaction type	Parity	Components	Gauge boson spin
Scalar	+	1	0
Pseudoscalar	-	1	0
Vector	-	4	1
Axial vector	+	4	1

Table 2.2: Different Lorentz-invariant currents are shown. Each of them has a unique combination of behavior under the action of the parity operator and the spin of the exchange particle of the interaction. Information merged from [10] chap. 11.2, tab. 11.1 and chap. 11.3, tab. 11.2.

Electroweak Unification

Particle physicists were always seeking for a unified theory describing all fundamental particles and their interactions. A large success in this desire of unification was the development of a theory by Glashow, Salam and Weinberg (GSW) in the 1960's which combined electromagnetism and weak interaction into a single picture for the first time and was awarded with the Nobel Prize in 1979 [15]. The charged-current weak interaction is invariant under $\mathrm{SU}(2)_L$ local phase transformations, where L

[4] The action of the parity operator $\hat{P}$ causes spatial inversion with respect to the origin of the coordinate system, i.e. $\vec{x} \to -\vec{x}$.

[5] The left- and right-handed chiral projection operators $P_R = \frac{1}{2}(1 + \gamma^5)$ and $P_L = \frac{1}{2}(1 - \gamma^5)$ allow for the decomposition of any Dirac spinor into a left- and a right-handed part. Here, the γ^5-matrix is given by $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$ where γ^j with $j = 0, 1, 2, 3$ denotes the Dirac matrices [10].

denotes the action on only left-handed particles as explained at the end of the last section. In order to fulfill this local gauge invariance, three gauge fields need to be introduced, corresponding to the gauge bosons $W^{(1)}$, $W^{(2)}$ and $W^{(3)}$. The physical W bosons are then given by the linear combinations [10]:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}\left(W_\mu^{(1)} \mp W_\mu^{(2)}\right) \quad . \tag{2.6}$$

Since experiments have proven that the neutral Z boson couples to left- and right-handed chiral particle states, matching the Z boson with the remaining $W^{(3)}$ field, which might be an obvious solution, is not correct. Instead, with the Z boson and the photon both being particles with an electric charge of zero, the U(1) gauge symmetry of the electromagnetic interaction can be replaced with a new $\mathrm{U}(1)_Y$ local gauge symmetry. $\mathrm{U}(1)_Y$ then gives rise to a new gauge field B_μ which again couples to a new kind of charge, called weak hypercharge Y. It is defined by [10]:

$$Y = 2\left(Q - I_W^{(3)}\right) \tag{2.7}$$

where Q is the electric charge and $I_W^{(3)}$ the third component of the weak isospin. The photon field A_μ and the Z boson can then be expressed as linear combinations of $W_\mu^{(3)}$ and B_μ with a weak mixing angle θ_W [10]:

$$A_\mu = +B_\mu \cos(\theta_W) + W_\mu^{(3)} \sin(\theta_W) \tag{2.8}$$

$$Z_\mu = -B_\mu \sin(\theta_W) + W_\mu^{(3)} \cos(\theta_W) \quad . \tag{2.9}$$

θ_W further links the electromagnetic coupling α_{em} with the weak one α_W by [10]

$$\sin^2(\theta_W) = \frac{\alpha_{\mathrm{em}}}{\alpha_W} = \frac{e^2}{g_W^2} \approx 0.23 \tag{2.10}$$

where g_W denotes the weak coupling constant. Measurements of $\sin^2(\theta_W)$ yield an average result of $\sin^2(\theta_W) = 0.23146 \pm 0.00012$ which is in good agreement with the theoretical prediction [10].

2.3 Quantum Chromodynamics

The quantum field theory of the strong interaction is called Quantum Chromodynamics. It is non-abelian, invariant under SU(3) local gauge transformations and introduces a new quantum number named "color", which is always conserved. The generators of the symmetry group have the form of 3×3 matrices, requiring three color states (red, green, blue) and the respective anti-colors as well [10]:

$$r = \begin{pmatrix}1\\0\\0\end{pmatrix}, \qquad g = \begin{pmatrix}0\\1\\0\end{pmatrix}, \qquad b = \begin{pmatrix}0\\0\\1\end{pmatrix} \quad . \tag{2.11}$$

Overall, there are eight generators which can be identified with the eight physical gluons already mentioned in table 2.1. Carrying different combinations of color and anti-color, they form an octet of

colored states [10]:

$$r\bar{g},\ \ g\bar{r},\ \ r\bar{b},\ \ b\bar{r},\ \ g\bar{b},\ \ b\bar{g},\ \ \frac{1}{\sqrt{2}}(r\bar{r}-g\bar{g}),\ \ \frac{1}{\sqrt{6}}(r\bar{r}+g\bar{g}-2b\bar{b})$$

As a result, only color-charged particles can couple to gluons and participate in the strong interaction. Thus, this behavior also allows for gluon self-interactions (in contrast to QED where two photons can not couple to each other). The naive expectation would now be to observe single quarks and gluons in nature which then carry color and, in case of the quarks, fractional electric charge. But this has never been the case yet and is explained by the (not analytically proven) hypothesis of *color confinement* which was already shortly mentioned in section 2.1. The attractive interactions between the virtual gluons, which are exchanged between two bound quarks, lead to a squeezing of the color field to the form of a tube. The energy stored in this field rises with the distance between the quarks in a linear way, making it impossible to separate them. Instead, this energy is used to create new quark-antiquark pairs from the vacuum which then rearrange with the initial quarks and form colorless hadrons again. *Color confinement* further limits the composition of bound hadronic states to either $q\bar{q}$ (mesons), qqq (baryons) or $\bar{q}\bar{q}\bar{q}$ (antibaryons). Combinations of $q\bar{q}$ and qqq (called pentaquarks) are theoretically allowed, but have not been observed yet [10].
Another feature making QCD unique compared the electromagnetic, weak or even unified electroweak theory is its coupling constant α_S which varies with the energy scale. This behavior splits QCD into two parts which have to be treated differently. In the regime of low momentum transfer, $|q| \approx 1\,\text{GeV}$, perturbation theory can not be used (since $\alpha_S = \mathcal{O}(1)$ is large) and other techniques have to be explored. The second part, typically referred to as *asymptotic freedom*, is reached for $|q| > 100\,\text{GeV}$ with α_S decreasing to ≈ 0.1 and allows for the application of perturbation theory again [10].

2.4 The Z Boson

2.4.1 Properties and Decay

The discovery of the Z boson by the experiments UA1 and UA2 at CERN in 1983 [16][17] in the decay channels $Z \to e^+e^-$ and $Z \to \mu^+\mu^-$ successfully confirmed the predictions made by Glashow, Salam and Weinberg years before. With a mass of $m_Z = (91.1876 \pm 0.0021)\,\text{GeV}$ and a width of $\Gamma_Z = (2.4952 \pm 0.0023)\,\text{GeV}$ [18] it can decay into various combinations of fermions and antifermions (except for top quarks since $m_t > m_Z$) which are listed in table 2.3. The large difference between the hadronic and leptonic branching fractions is caused by the color charge carried by the quarks, allowing for a decay into an either $r\bar{r}$, $g\bar{g}$ or $b\bar{b}$ quark-antiquark pair. In contrast to the weak charged-current interaction, which is mediated by the W boson, the Z, representing the weak neutral current, couples to left- and right-handed chiral particle states, although the couplings are not equal. This effect is due to the mixing of the B_μ and $W_\mu^{(3)}$ fields where the first one comes from the $\text{U}(1)_Y$ symmetry of the electromagnetic interaction whose coupling is not affected by the particles chirality. The couplings of the Z to left- and right-handed chiral states are given by [10]

$$c_L = I_W^{(3)} - Q\sin^2(\theta_W) \quad \text{and} \quad c_R = -Q\sin^2(\theta_W) \tag{2.12}$$

with Q being the charge of the fermion the Z couples to.
Consequently, the Z and the photon both couple to all flavours of fermions and one can replace

Decay mode	$\mathcal{BR}$
e^+e^-	(3.3632 ± 0.0042) %
$\mu^+\mu^-$	(3.3662 ± 0.0066) %
$\tau^+\tau^-$	(3.3696 ± 0.0083) %
$\nu_{e,\mu,\tau}\bar{\nu}_{e,\mu,\tau}$	(20.000 ± 0.055) %
Hadrons	(69.911 ± 0.056) %

Table 2.3: The different decay modes of the Z boson and their branching ratios. The branching ratio $\mathcal{BR}$ is defined as the partial decay width Γ_j into a certain final state j divided by the total width Γ_Z of the Z boson. Data from [18].

the γ in each QED Feynman diagram with a Z. Nevertheless, their different masses result in an energy-dependence of the matrix element for such a process. This behavior has been used at the LEP collider at CERN in the 1990's for extensive studies of the Z boson and its properties. Considering the s-channel annihilation process $e^+e^- \to Z \to q\bar{q}$, the whole center-of-mass energy $\sqrt{s}$ is transferred to the intermediate state particle Z/γ. While for $\sqrt{s} \ll m_Z$ the QED process mediated by the photon dominates, the couplings to γ and Z become similar for $\sqrt{s} \gg m_Z$, resulting in both processes needed to be taken into account. In the range between ($\sqrt{s} \approx m_Z$) the interaction is dominated by the Z boson, as shown in figure 2.3. The contribution of the pure $e^+e^- \to Z \to q\bar{q}$ process to the measured cross section σ then follows a so-called Breit-Wigner distribution which is given by [10]:

$$\sigma \propto \frac{1}{\left(s - m_Z^2\right)^2 + m_Z^2\Gamma_Z^2} \quad . \tag{2.13}$$

In addition, the measurement of the Z resonance provides another interesting possibility of testing the Standard Model predictions, namely the determination of the number N_ν of neutrino families which can directly be obtained from the full width at half maximum (FWHM = Γ_Z) of the resonance peak [10]:

$$N_\nu = \frac{\Gamma_Z - 3\Gamma_{ll} - \Gamma_{\text{hadrons}}}{\Gamma_{\nu\nu}^{\text{SM}}} \quad . \tag{2.14}$$

Here, Γ_{ll} is the partial decay width into a single flavor of charged leptons, Γ_{hadrons} the decay width into all quark final states and $\Gamma_{\nu\nu}^{\text{SM}} = 167\,\text{MeV}$ the prediction of the Standard Model for the decay width of the Z boson into a $\nu\bar{\nu}$ pair. From the recorded data, one finds $N_\nu = 2.9840 \pm 0.0082$ which is in good agreement with the SM prediction of $N_\nu^{\text{SM}} = 3$ [10].

2.4.2 ZZ Diboson Production

At the *Large Hadron Collider* (LHC), which is the successor of the LEP accelerator at CERN and is introduced in chapter 3, Z bosons can also be produced in pairs. Measurements of this diboson production allow for further testing of many SM predictions made by electroweak and QCD theories since any major deviations would give a hint to beyond Standard Model physics [19]. The production cross section has been measured by the ATLAS experiment (cf. sec. 3.2) at $\sqrt{s} = 8\,\text{TeV}$ to be

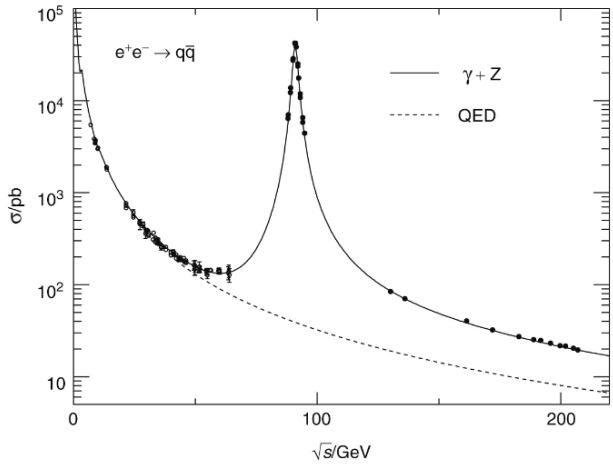


Figure 2.3: Cross section measurements in the $e^+e^- \to q\bar{q}$ channel in the energy region near the Z resonance performed by LEP and earlier experiments. In addition, the pure QED contribution is visualized by a dashed line. Image from [10].

$7.1^{+0.5}_{-0.4}$ (stat.) ± 0.3 (syst.) ± 0.2 (lumi.) pb [20]. From figure 2.4 it can be found that this value is of about four orders of magnitude lower than the total cross section for single Z boson production. In the SM, ZZ production can either occur quark- or gluon-initiated, where the latter mode is suppressed compared to the first one due to its loop-structure as shown in the Feynman diagrams in figure 2.5. Some BSM theories also predict the existence of triple gauge couplings which could give rise to another ZZ boson production channel as visualized in figure 2.6 and would increase the cross section in regions of high energy scale interactions. Such couplings are forbidden in the Standard Model since they would violate the $SU(2)_L \times U(1)_Y$ symmetry of the unified electroweak interaction introduced in section 2.2 [19].

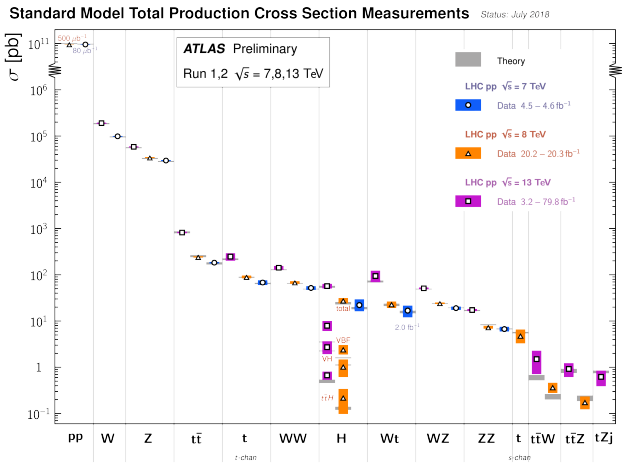
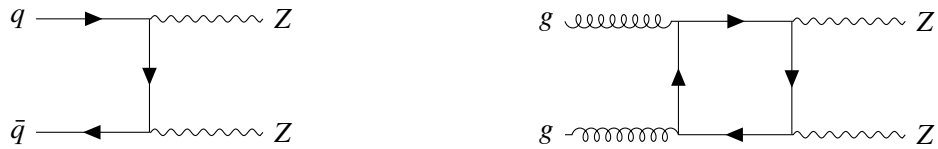


Figure 2.4: Standard Model total production cross section measurements of different particle states involving gauge bosons and top quarks done by the ATLAS experiment during run 1 and 2 of the LHC. Image from [21].



(a) Quark-initiated (b) Gluon-initiated

Figure 2.5: Leading-order Feynman diagrams for SM ZZ production in pp-collisions [14].

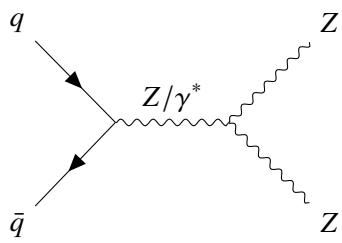


Figure 2.6: Example Feynman diagram for BSM ZZ production including a triple gauge coupling [14].

2.5 Higgs Mechanism

2.5.1 Electroweak Symmetry Breaking and the Higgs Field

In the Standard Model, the fundamental fermions and massive gauge bosons gain their masses by an interaction with the so-called Higgs field and a spontaneous breaking of the $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ electroweak symmetry. Particle states are then described by excitations of this field. Nowadays, this theory is known as the Higgs mechanism [22] [23].
In order to generate the masses of the three massive electroweak gauge bosons, the minimal SM Higgs model requires four degrees of freedom which are provided by two complex scalar fields. These further need to account for the charges of the gauge bosons, resulting in one field being charged while the other one is neutral [10]:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \quad . \tag{2.15}$$

The respective Lagrangian is then given by [10]

$$\mathcal{L} = (\partial_\mu \phi)^\dagger (\partial^\mu \phi) - V(\phi) \tag{2.16}$$

where the Higgs potential $V(\phi)$ reads [10]:

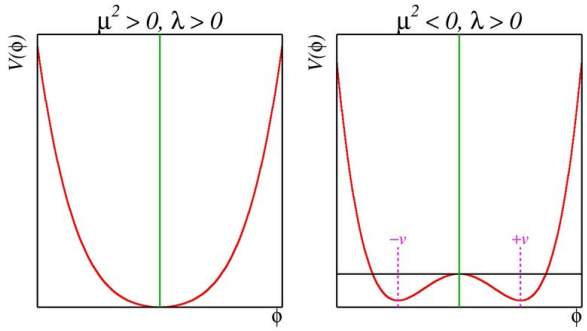
$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 \quad , \quad \lambda > 0 \ . \tag{2.17}$$

Demanding the parameter λ to be positive ensures that the potential has a finite minimum, but its position and the potential form further depend on μ^2. While the minimum occurs at $\phi = 0$ for positive

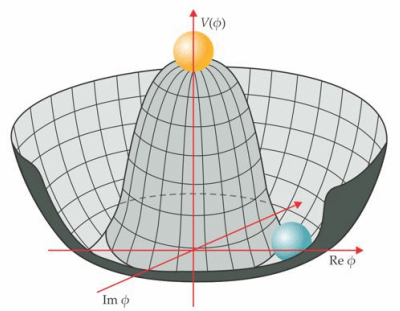
values of μ^2, there are two of them at the positions [10]

$$\phi = \pm v = \pm \left| \sqrt{\frac{-\mu^2}{\lambda}} \right| \tag{2.18}$$

for $\mu^2 < 0$, as visualized in figure 2.7. This behavior results in a vacuum state of the field of either $\phi = +v$ or $\phi = -v$ unequal to zero and is called *spontaneous symmetry breaking* [10].



(a) Two-dimensional sketch of the potential for $\mu^2 > 0$ (left) and $\mu^2 < 0$ (right). The position of the minimum or minima, respectively, depends on the sign of μ^2. Image from [24].



(b) Three-dimensional sketch of the potential for $\mu^2 < 0$. The position of the ground state is marked by the blue ball while the orange one denotes the origin of the coordinate system in Re(ϕ) and Im(ϕ). Image from [25].

Figure 2.7: The potential $V(\phi) = \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2$ for a complex scalar field ϕ.

The masses of the gauge bosons can then be calculated from the first term in equation 2.16, but first the partial derivative ∂_μ needs to be replaced with an appropriate covariant derivative D_μ to ensure to not violate the $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ local gauge symmetry of the electroweak model [10]:

$$\partial_\mu \;\rightarrow D_\mu = \partial_\mu + i g_W \vec{T} \cdot \vec{W}_\mu + i g' \frac{Y}{2} B_\mu \;. \tag{2.19}$$

Here, g_W, W_μ, Y and B_μ are defined as introduced in section 2.2 and g' denotes the coupling constant of the electromagnetic $\mathrm{U}(1)_Y$ symmetry. Further is $\vec{T} = \frac{1}{2}\vec{\sigma}$ where $\vec{\sigma}$ is a vector composed of the Pauli matrices σ_1, σ_2 and σ_3. Applying this covariant derivative to equation (2.16) and calculating $(D_\mu\phi)^\dagger(D^\mu\phi)$ (see [10] for details) finally yields the gauge boson masses of [10]:

$$m_W = \frac{1}{2} g_W v \quad , \quad m_A = 0 \quad \text{and} \quad m_Z = \frac{1}{2} v \sqrt{g_W^2 + (g')^2} \;. \tag{2.20}$$

Thus, the W and Z boson acquire mass through the Higgs mechanism while the photon remains massless. The couplings of the gauge bosons to the Higgs field scale linearly with the boson mass, i.e. $\propto m_W$ and $\propto m_Z$ for the massive bosons, respectively, and zero for the photon. The vacuum expectation value v of the Higgs field can directly be obtained from equation (2.20) and yields a value of [10]:

$$v = 246\,\mathrm{GeV} \;. \tag{2.21}$$

Beside the gauge bosons, also the fermions of the Standard Model acquire their mass through the Higgs mechanism. Their so-called Yukawa couplings g_f to the Higgs field are given by [10]

$$g_f = \sqrt{2}\frac{m_f}{v} \tag{2.22}$$

where m_f denotes the mass of the fermion. As a consequence, higher mass particles couple more strongly to the Higgs field than lighter ones do.

2.5.2 The Higgs Boson

The Higgs boson h physically occurs as an excitation of the Higgs field and is a neutral scalar spin-0 particle. Its mass is predicted by the Standard Model as $m_h = \sqrt{2\lambda v^2}$ and has been measured to be (125.10 ± 0.14) GeV [26]. At the LHC, Higgs bosons are mainly produced by either gluon-gluon fusion (ggF) or vector boson fusion (VBF), as shown in figure 2.8. Given that the Higgs boson's coupling to the gauge bosons and fermions increases with the mass of the respective particle, the Higgs always prefers to decay into the most massive particles that are accessible while taking energy conservation into account. The mass-dependent branching ratios of the Higgs boson are depicted in figure 2.9 and, for a mass of $m_h = 125$ GeV, summarized in table 2.4. Decays into WW or ZZ are only possible if one gauge boson is produced virtual, i.e. off-shell, which is marked with a little star behind the particle. Although photons and gluons do not have any mass, the Higgs boson can decay into such final states through loops of virtual top quarks or W bosons [10].

(a) Gluon-gluon fusion

(b) Vector boson fusion

Figure 2.8: The two dominant Higgs boson production modes at the LHC. Gluon-gluon fusion has a much larger cross section, while vector boson fusion allows for an easier detection due to the signature of two forward jets [14].

2.6 The τ-Lepton

The τ-lepton is the third generation partner of the electron and concurrently the heaviest lepton. Due to a high mass of $m_\tau = (1.77682 \pm 0.00016)$ GeV it decays rather quickly and only has a short lifetime of $\tau_\tau = (2.903 \pm 0.005) \times 10^{-13}$ s [27]. In contrast to electrons or muons, its mass further allows the tau to either decay into lighter leptons or into different combinations of charged and neutral pions. The dominant decay channels are depicted in figure 2.10. In this analysis, hh and ZZ decays into four τ-leptons will be considered, where either all four or three taus further decay hadronically and no or one tau leptonically, respectively. The full hadronic decay mode is then referred to as $4h0l$ while the channel including a light lepton is denoted with $3h1l$.

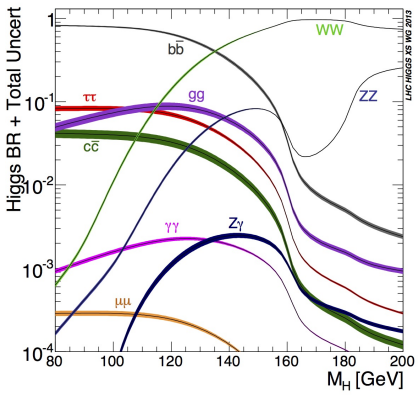


Figure 2.9: Branching ratios of the Higgs boson into possible final states depending on its mass M_H. The uncertainty of the branching ratio is visualized by the thickness of the respective band. Image from [28].

Decay mode	$\mathcal{BR}$
$b\bar{b}$	57.8 %
WW^*	21.6 %
gg	8.6 %
$\tau^+\tau^-$	6.4 %
$c\bar{c}$	2.9 %
ZZ^*	2.7 %
$\gamma\gamma$	0.2 %

Table 2.4: The dominant decay modes of the Standard Model Higgs boson and their branching ratios. Data from [10].

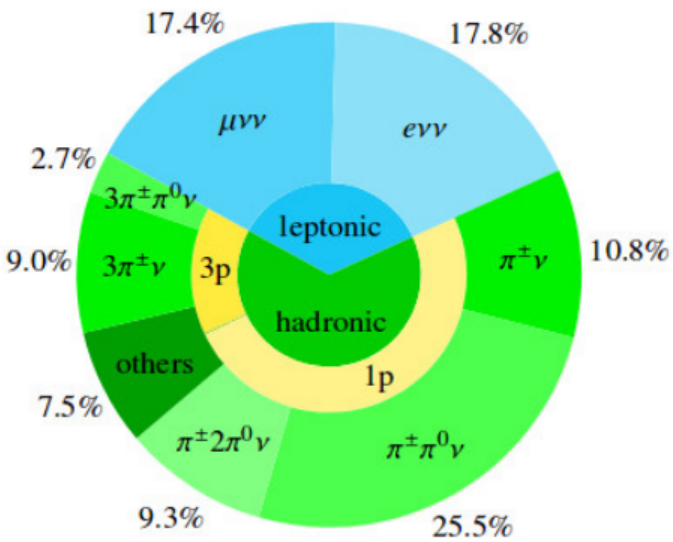


Figure 2.10: The dominant decay modes of the τ-lepton and their branching ratios. Tau decays are separated into final states containing charged leptons (*leptonic*) and those without (*hadronic*). The hadronic decays are further discriminated by the number of charged pions in the decay products, referred to as *prong* (p). Image from [29].

CHAPTER 3

Experimental Setup

Since our universe is only built from the Standard Model's first generation charged particles, other fermions or gauge bosons have to be created artificially if their properties should be studied. For many decades now, physicists use particle accelerators for that purpose. Today, they either collide two particle beams or one beam with a fixed target and use large detector instruments to precisely measure the particles created in these collisions. For a fixed beam energy, the collision of two beams results in a higher center-of-mass energy available for particle creation than the fixed target setup, making it the favorable design of most currently operated high energy accelerators.

3.1 The Large Hadron Collider

The *Large Hadron Collider* (LHC) is a circular particle accelerator with a circumference of approximately 27 km, located at the European Organization for Nuclear Research (CERN) near Geneva, Switzerland. It accelerates protons which then collide with a center-of-mass energy of currently $\sqrt{s} = 13\,\text{TeV}$. Additionally, there are special data taking periods with heavy ion beams. Along the accelerator, four large detectors are placed in order to study the properties of the particles created in these collisions and to search for new particles predicted in theories and models beyond the Standard Model. The structure of the LHC including its pre-accelerators is depicted in figure 3.1.
After achieving an energy of 450 GeV in the Super Proton Synchrotron (SPS), the protons are injected in bunches into the LHC. One bunch contains a total of 1.2×10^{11} protons while the beam consists of 2808 bunches with a time spacing of 25 ns. This results in an instantaneous luminosity of $\mathcal{L} = 1.2 \times 10^{34}\,\text{cm}^{-2}\,\text{s}^{-1}$ [30] (for the year 2017) which has been increased over the years of operation multiple times and should also be further enhanced in the future (High-Luminosity LHC, [31]). The luminosity is an important parameter for such collider experiments since it is directly linked to the number of events N occurring of a certain physical process with cross section σ [11]:

$$N = \sigma \cdot \mathcal{L}_{\text{int}} = \sigma \cdot \int \mathcal{L}\,\mathrm{d}t \ . \tag{3.1}$$

The accelerator ring is built up of two almost circular vacuum tubes where the two protons beams are accelerated to their maximum energy of 6.5 TeV before they circulate for a few hours in the ring. Since the LHC is a synchrotron, the beam particles only gain energy when passing one of at total eight radio frequency cavities per turn. In the remaining revolution time, they are bent in magnetic

fields created by 1232 large main dipole and 392 quadrupole magnets in order to stay at the desired circular orbit. In such an accelerator, the energy loss per turn due to synchrotron radiation decreases with higher particle mass and hence is much lower for protons or heavy ions than e.g. for electrons. Therefore, massive particles are a more suitable choice when high center-of-mass energies should be reached at which many particle physicists hope to find hints for new physics beyond the theory of the Standard Model [30] [32].

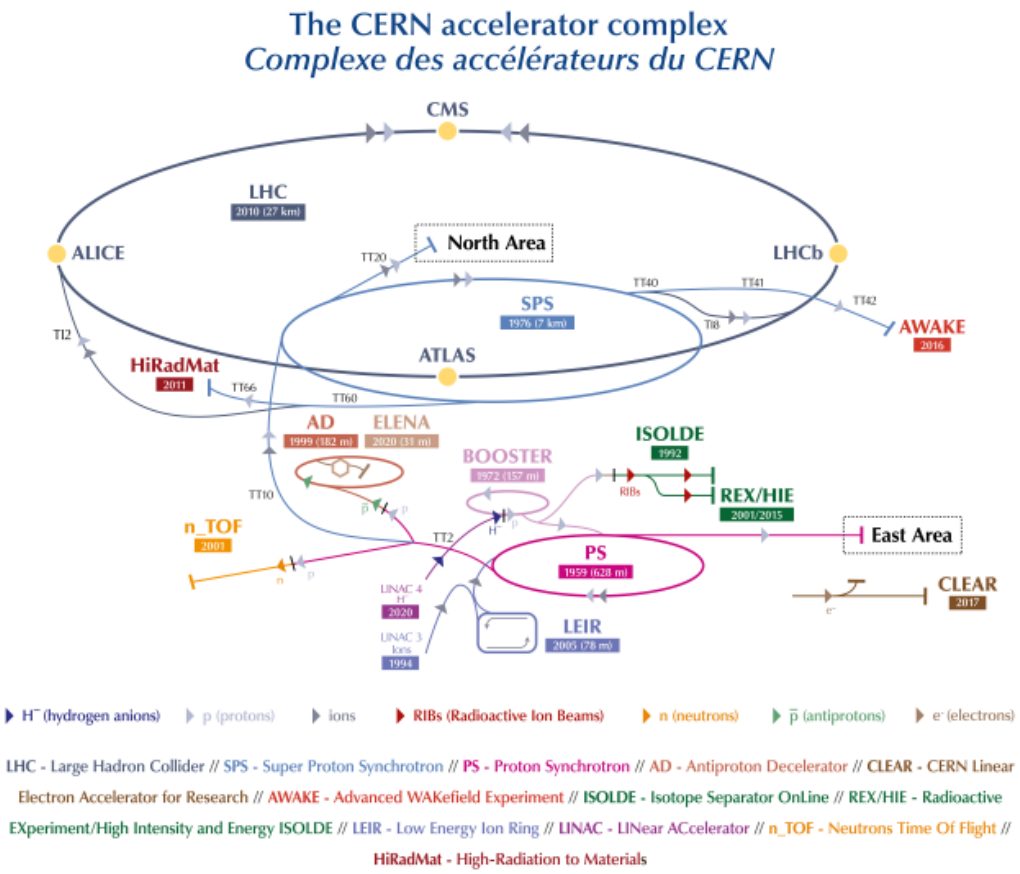


Figure 3.1: Setup of the accelerator complex at CERN including the different experiments. Protons gain energy in four pre-accelerators (LINAC 4, BOOSTER, PS, SPS) before being injected into the LHC. Heavy ions use LINAC 3 and LEIR instead, followed by PS and SPS as for protons. Image from [33].

3.2 The ATLAS Detector

The ATLAS (**A** **T**orodial **L**HC **A**pparatu**S**) detector [34] is the largest detector which has ever been built for a particle accelerator. It has a forward–backward symmetric cylindrical form with a diameter of 25 m, a length of 44 m, weighs 7 000 t and is built as a multipurpose detector. As shown in figure 3.2, its subsystems are wrapped around the interaction point at the center and cover almost the entire solid angle. The different components and the used coordinate system are explained in more detail in

the following subsections [35].

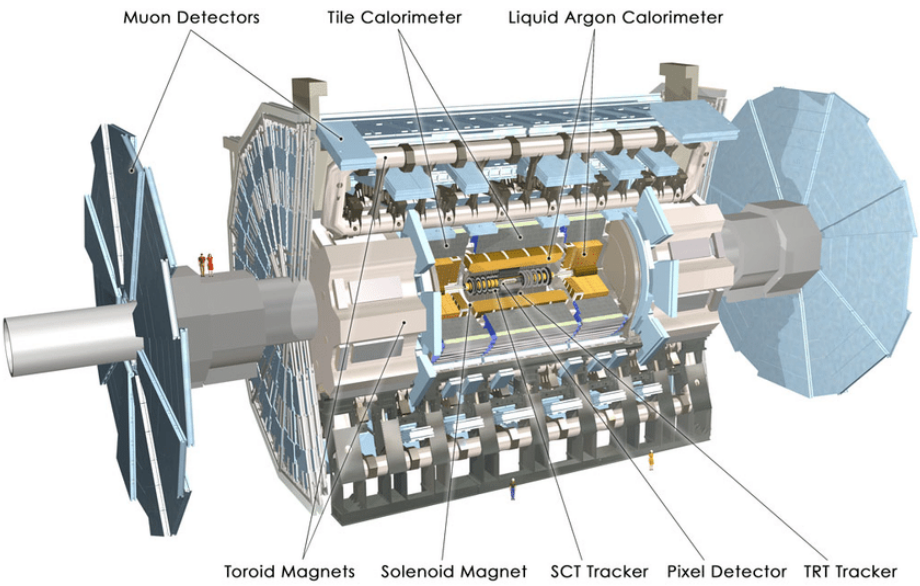


Figure 3.2: The ATLAS detector: the pixel detectors and other systems for reconstruction of the particle tracks are located at the center, surrounded by a homogeneous magnetic field created by superconducting magnets which bend the particles and therefore allow for the measurement of their momentum. The adjacent electromagnetic and hadronic calorimeters measure the energy of the particles through their interaction with the detector material. The muon systems at the outer end complete the detector, they measure precisely the tracks of created muons which do not deposit any significant amount of energy in the calorimeters. Image from [36].

3.2.1 ATLAS Coordinate System

The ATLAS coordinate system is depicted in figure 3.3. Its origin lies exactly at the position of the nominal collision point. The particle beams are moving in $+z$ and $-z$ direction, while x- and y-axis define the transverse detector plane with positive x showing into the direction of the center of the LHC and positive y pointing upwards. In addition, the polar angle θ is given as the angle between particle track and beam axis whereas the azimuthal angle ϕ is measured in the transverse plane around the beam axis [34]. Defining the pseudorapidity as

$$\eta = -\ln\tan(\theta/2) \tag{3.2}$$

the three-dimensional distance measure can be described by the observable ΔR:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \ . \tag{3.3}$$

Since particles created in the collision can carry an arbitrary amount of momentum and energy in z-direction, only their transverse components p_{T} and E_T are of interest for the measurement process. Assuming natural units with $\hbar = c = 1$, both quantities describe the same information and usually

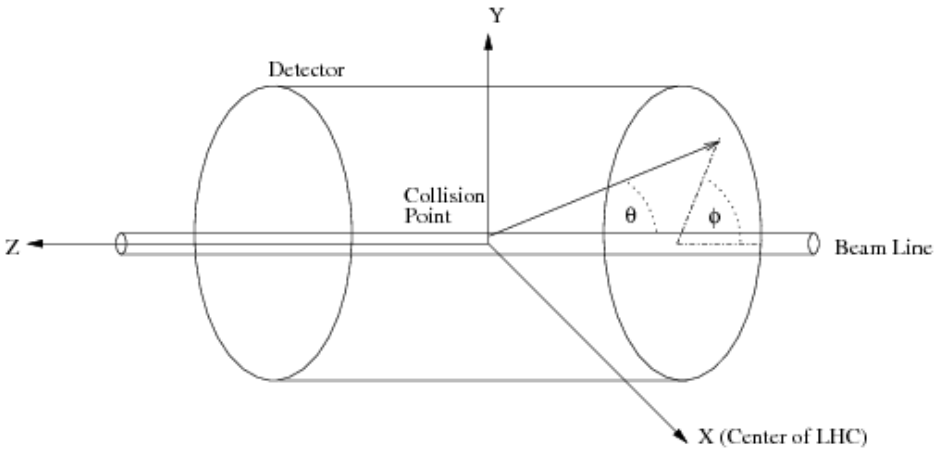


Figure 3.3: Coordinate system of the ATLAS detector. The positive x-axis points to the center of the LHC, the positive y-axis upwards and the z-axis follows the direction of the beam line. The azimuthal angle in the transverse plane around the z-axis is given by ϕ while θ describes the polar angle between the particle track and the beam line. Image from [38].

only p_{T} is used as event observable which is then given by [34]:

$$p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2} \ . \tag{3.4}$$

With the initial momentum of the colliding particles in x- and y-direction being zero, one can then make use of momentum conservation and define the missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$ as the magnitude of the negative vector sum of the momenta of all particles detected in a collision [37]:

$$E_{\mathrm{T}}^{\mathrm{miss}} = \left| -\sum_i p_{T,i} \right| \ . \tag{3.5}$$

3.2.2 Triggers

In order to be able to select interesting events of a variety of physics processes, a highly efficient triggering system with sufficient background rejection is required, especially for objects with low transverse momentum p_{T}. At the design luminosity of $10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, proton-proton collisions occur with a rate of approximately 1 GHz but the current available technology limits the data recording rate to a value close to 200 Hz. The required rejection factor of 5×10^6 can easily be calculated and is achieved by a three-level trigger system. The Level-1 trigger (L1) decreases the data rate to approximately 75 kHz by using a subset of the whole detector information. Basing on that it decides to either further process the event or to reject it. After that follow the Level-2 (L2) trigger and the event filter which finally reduce the data rate to about 200 Hz and usually are called the high-level trigger [34].

3.2.3 Tracking & Vertexing in the Inner Detector

The inner detector (ID) is the measurement instrument closest to the collision point and therefore needs to deal with a huge number of particle tracks, namely approximately 1000 every 25 ns within the accessible pseudorapidity range of $\eta < 2.5$. A combination of pixel and silicon microstrip (SCT) trackers embedded in a 2 T axial magnetic field followed by straw tubes in the Transition Radiation Tracker (TRT) offers the necessary fine detector granularity which allows for high-precision measurements of particle momenta and vertex positions [34]. It is depicted in figure 3.4.

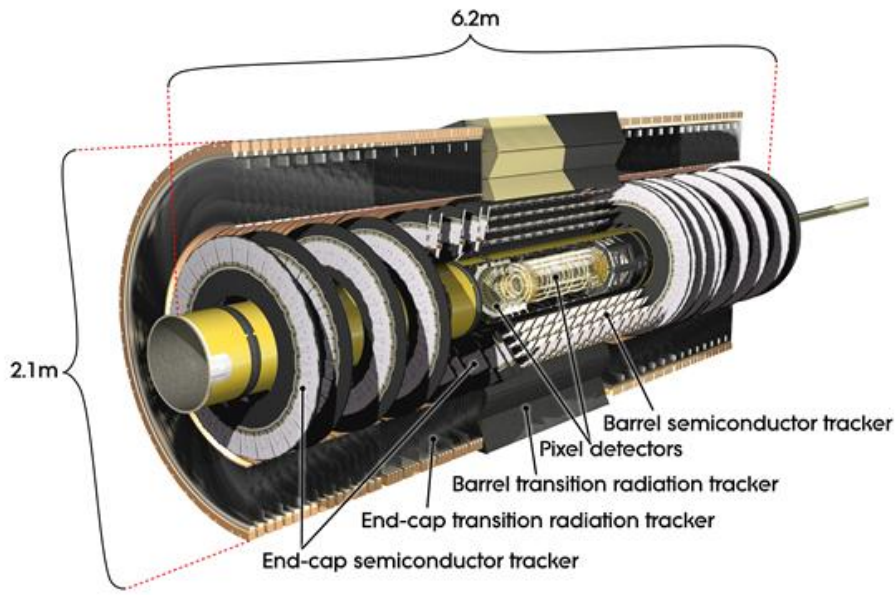


Figure 3.4: Overview of the ATLAS inner detector (ID). It consists of three major components, starting with the silicon pixel detectors closest to the interaction point, followed by the semiconductor tracker (SCT) and the Transition Radiation Tracker (TRT). Image from [39].

The pixel detector has approximately 80.4×10^6 readout channels and each track usually crosses three of its layers which are segmented in $R - \phi$ and z. Each pixel sensor has a minimum size of $50 \times 400\,\mu\text{m}^2$ in $R - \phi \times z$ achieving an intrinsic precision of 10 µm ($R - \phi$) and 115 µm in z (barrel) or R (end-cap), respectively [34].
In the subsequent SCT, each track hits eight strip layers corresponding to four space points. The barrel region is equipped with 40 mrad stereo strips which measure R and ϕ as well as the difference of both quantities with one set of strips in each layer parallel to the z-axis. For the end-cap, the stereo strips are supplemented by a radially running set of strips instead. The average strip pitch is about 80 µm and the intrinsic precision achieves values of 17 µm in $R - \phi$ and 580 µm in z (barrel) or R (end-cap), respectively. Overall, the SCT consists of approximately 6.3×10^6 readout channels [34].
The TRT offers only a measurement of $R - \phi$, but therefore each track typically leaves 36 hits in the detector which allows for tracking up to $|\eta| < 2.0$. Its straw tubes with a diameter of 4 mm provide an intrinsic precision of 130 µm per straw. The TRT has 351 000 readout channels [34].
Built in this way, the inner detector allows for high-precision measurements of $R - \phi$ and z of particle tracks. The first subsystem further enables impact parameter determination and tagging of heavy-flavor

quark jets as well as of τ-leptons. The momentum measurement is performed by both the pixel detectors and the TRT straw tubes where in the latter the lower precision per point is compensated by an overall larger amount of measurements and measured track length. Furthermore, the TRT provides the possibility of electron identification by the detection of transition radiation photons [34].

3.2.4 Calorimetry

The energy which a particle looses due to interaction with the detector material is measured in the calorimeters. If the particle gets completely stopped in the material, all of its energy is deposited in the calorimeter clusters. The ATLAS detector contains an electromagnetic calorimeter built of lead and liquid argon (LAr) and a hadronic one which are both separated into a barrel and an end-cap part again and cover the range up to $|\eta| < 4.9$. They are illustrated in figure 3.5 [34].

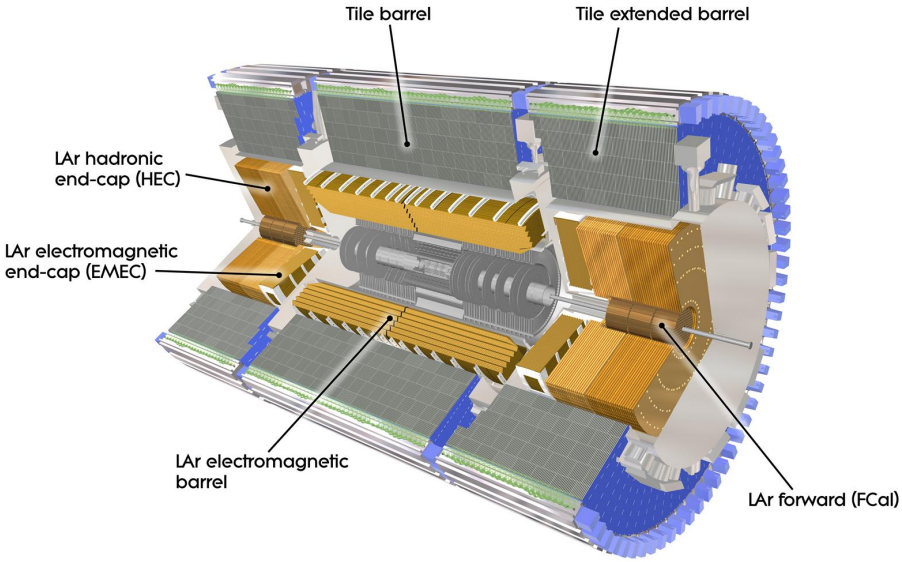


Figure 3.5: Overview of the ATLAS calorimeters. The electromagnetic barrel and both calorimeter end-caps work with liquid argon (LAr). An also LAr based forward calorimeter, the hadronic tile barrel and the tile extended barrels complete the detector design. Image from [40].

The inner region of the electromagnetic calorimeter, which matches the η-coverage of the inner detector, provides excellent energy determination of electrons and photons creating showers by interaction with the detector material. The outer part has a coarser granularity but still allows for adequate jet reconstruction and measurements of missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$ [34].
Hadronic calorimetry of strongly interacting particles is performed in overall three different calorimeters. The tile calorimeter made of steel and scintillating tiles surrounds the electromagnetic barrel and is segmented in three layers. Including the extended barrel, it covers an η-region up to $|\eta| < 1.7$. Two separate wheels per side built the hadronic end-caps which are directly placed behind the electromagnetic end-cap calorimeters, using liquid argon in combination with copper plates. They have a small overlap with the tile as well as with the forward calorimeter in order to decrease the

material gap in the transition area between the components. The forward calorimeter finally also covers the regions along the the beam axis and further reduces the radiation background in the adjacent muon spectrometers [34].

3.2.5 Muon System

Since muons barely interact with the detector material, they deposit no significant amount of energy in the calorimeters. In addition to the measurement performed in the inner detector, their tracks are measured again in over 5 000 muon chambers at the outer parts of the ATLAS detector. Those are equipped with large superconducting toroid magnets, separate trigger instruments and high-precision tracking chambers covering a region of up to $|\eta| < 2.7$. Due to their large size, they are well visible in figure 3.2. The measurement of particle coordinates and track curvature is mainly performed by 1 150 Monitored Drift Tubes and 32 additional Cathode Strip Chambers with higher granularity in the outer η-region. Triggering is done with Resistive Plate Chambers in the the barrel and Thin Gap Chambers in the end-caps for a region of up to $|\eta| < 2.4$ [34].

3.3 Object Reconstruction at ATLAS

Identifying the different particles created in a proton-proton collisions at the LHC is a key requirement for the further analysis of all physical processes taking place in the collision. Hence, there is a need for dedicated reconstruction and identification algorithms for a variety of particle types. In the following subsections, a short overview of the used methods for jets (3.3.1), light leptons and photons (3.3.2) and missing transverse energy (3.3.3) is given. Identification of hadronically decaying τ-leptons in ATLAS is of special importance for this thesis and will therefore be treated separately in chapter 5.

3.3.1 Jets

Jets are collimated particle beams of color charged quarks and gluons which are created by the strong force in a collision and emerge from the interaction point in a huge amount. Since in the Standard Model free particles are not allowed to carry color charge, they recombine to color neutral hadrons after very short timescales again. An example of a two-jet event is depicted in figure 3.6. Hadrons in a jet usually have a comparably low momentum of approximately 300 MeV [41] perpendicular to the jet axis. The total jet energy is given by the energy of the quark from which the jet originates. In order to determine this energy, one thus needs to reconstruct as many of the created hadrons as possible and assign them to the jet [41].
In ATLAS, jet reconstruction starts with three-dimensional topological clusters which group together calorimeter cells. Those with a signal-to-noise ratio of at least 4σ serve as seed clusters which are then iteratively expanded over all adjacent cells with a minimum ratio of 2σ. At the end of this process, all direct neighboring cells are added to the cluster in addition. This searching procedure further includes a splitting step, in which all cells are checked for local energy content maxima. If such a maximum is found, the cluster is split into two separate ones for which the iteration process starts over again. The final criteria for a reconstructed cluster are an energy equal to the energy sum of its single cells, a mass of zero and the point from which the jet originates has to match the collision point of the initial proton beams. After calibration (see [42]), jets are then built using the anti-k_t algorithm [43] and further processed by taking pile-up, jet origin and jet energy corrections into account [42][44].

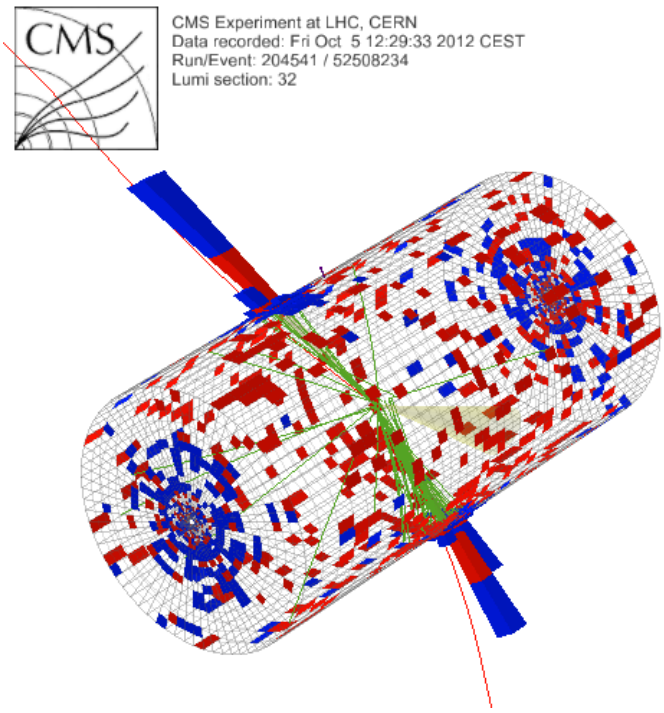


Figure 3.6: Typical two-jet event in the CMS detector resulting from a proton-proton collision. The particle beams enter the detector from bottom left and top right, respectively, and collide at the center, i.e. from where the green tracks emerge. The reconstructed jets are visible as red and blue cones crossing the detector plane. Image from [45].

3.3.2 Leptons & Photons

In contrast to jets and hadronically decaying τ-leptons, light leptons and photons can be reconstructed as isolated objects in the detector which makes their identification easier compared to objects with many tracks and cluster hits. Leptonically decaying taus ($\tau^- \to l^- \nu_\tau \bar{\nu}_l$ where $l = e, \mu$) are reconstructed as light leptons as well, since the neutrinos in the decay products do not interact with the detector material and therefore can not be detected directly.

Electrons & Photons

Reconstruction of electrons and photons in the central region ($|\eta| < 2.5$) is mainly based on information from shower shape variables in the electromagnetic calorimeter and on energy leakage in the hadronic calorimeter. The electron identification is further improved by making use of track quality variables and information about the matching between track and cluster position. For the forward region ($2.5 < |\eta| < 4.9$) which is not equipped with any tracking instruments, only cluster and shower shape information can be used for discrimination against hadrons. The number of defined working points[1] is

[1] Each fixed working point corresponds to a certain signal efficiency and background rejection factor. Both quantities are not independent from each other, i.e. increasing the signal efficiency usually results in a lower background rejection (and vice versa).

three for electrons (loose, medium, tight) while for photons only two (loose, tight) exist. Finally, the electromagnetic energy scale is further tuned using information from well known $Z \rightarrow ee$ decays [42].

Muons

Muons in the ATLAS detector are reconstructed by combination of the track and momentum information provided independently by the inner detector and the muon spectrometers. This method yields an efficiency of more than 95 % while the achieved momentum resolution is about 3 % over a wide p_{T} range and increases to 10 % at a p_{T} of 1 TeV [42].

3.3.3 Missing Transverse Energy

Missing transverse energy (MET) reconstruction again starts from topological clusters cells but also from identified muons. The same calibration method as for jets is used and gets further combined with the calibration of all physics objects in each final state. All clusters which could not be assigned to any reconstructed object are now collected, building the MET soft term. For a given event, the Soft Term Vertex Fraction (STVF) is then defined as the quotient of the p_{T} sum of all tracks from the hard-scatter vertex which could not be matched to jets and all tracks from all vertices which could not be matched to jets. Afterwards, the MET soft term is rescaled by this STVF, event-by-event. Furthermore, MET performance and relevant systematic uncertainties determined from data-simulation agreement of the MET distribution in leptonic W and Z decays are taken into account for reconstruction [42].

3.4 Event Generation & Monte Carlo Simulation

In order to allow for the comparison of recorded data with the theoretical predictions of the Standard Model, a good modeling of the underlying physics processes is required. For that purpose, Monte Carlo simulations are used which can be divided in two main steps. First, the physics process of interest is modeled by an event generator while afterwards a virtual detector algorithm simulates the interaction of the ATLAS detector with the created particles. The first part can further be split in three steps which are shortly summarized below and depicted in figure 3.7:

- The actual **hard collision** of two partons (one from the colliding protons each) with a large momentum transfer can be described by perturbation theory. Nevertheless, this part is often only simulated in leading order because higher order contributions are difficult to calculate [46].

- After the two-parton collision, so-called **parton showers** are created by processes of the strong interaction (gluon radiation, quark-antiquark creation from radiated gluons and more). They are simulated by an algorithm which scales, starting from the hard scattering process, the momentum transfer down to low energy scales, taking *color confinement* into account [46].

- The **hadronization** is the transition process from the partonic (color charged particles) to the hadronic final state in which only color neutral particles exist. Since lattice QCD, which is the only theoretical ansatz for the description of this transition, is formulated in Euclidean space-time and thus can not be applied to e.g. the time-evolution of partons into hadrons, only a few models for this step exist. The most prominent ones are the Lund string model [47] and the

cluster model (see [46] for more information). In the latter, all single gluons first annihilate into quark-antiquark pairs from which then color neutral clusters are formed [46].

For the generation of events used in this thesis, there are three relevant Monte Carlo generators, all starting from the initial proton-proton collision:

- **PYTHIA 8** [48] is a new C++ version of the PYTHIA general purpose generator including new physics models. Used as default generator in ATLAS, it starts with the hard scattering process which is calculated in lowest order QCD. Afterwards, further QED and QCD radiation in a shower approximation as well as underlying event activity is added. PYTHIA 8 offers a good mix of power and speed and is often used together with other generators for a variety of final states [49].

- **Sherpa** [50] is another multi purpose generator using parts of PYTHIA 8, but gives better performance for final state with a large number of isolated jets since it does not use a pure QCD showering algorithm as PYTHIA 8 does. In addition, it works with a further developed underlying event generation, based on the PYTHIA 8 version [49].

- **Powheg** [51] which is based on the MC@NLO generator [52] but produces only events with positive weight, is a more specialized generator for production of top quark events since it models their p_{T} distribution in a better way than e.g. PYTHIA 8 does. As stated by the name, it calculates the hard scattering processes at next to leading order in QCD perturbation theory [49].

Subsequently, the response of the ATLAS detector to these generated events is simulated. In order to do so, a copy of the detector is built in GEANT4 [53] including the whole detector configuration, misalignments and other distortions.

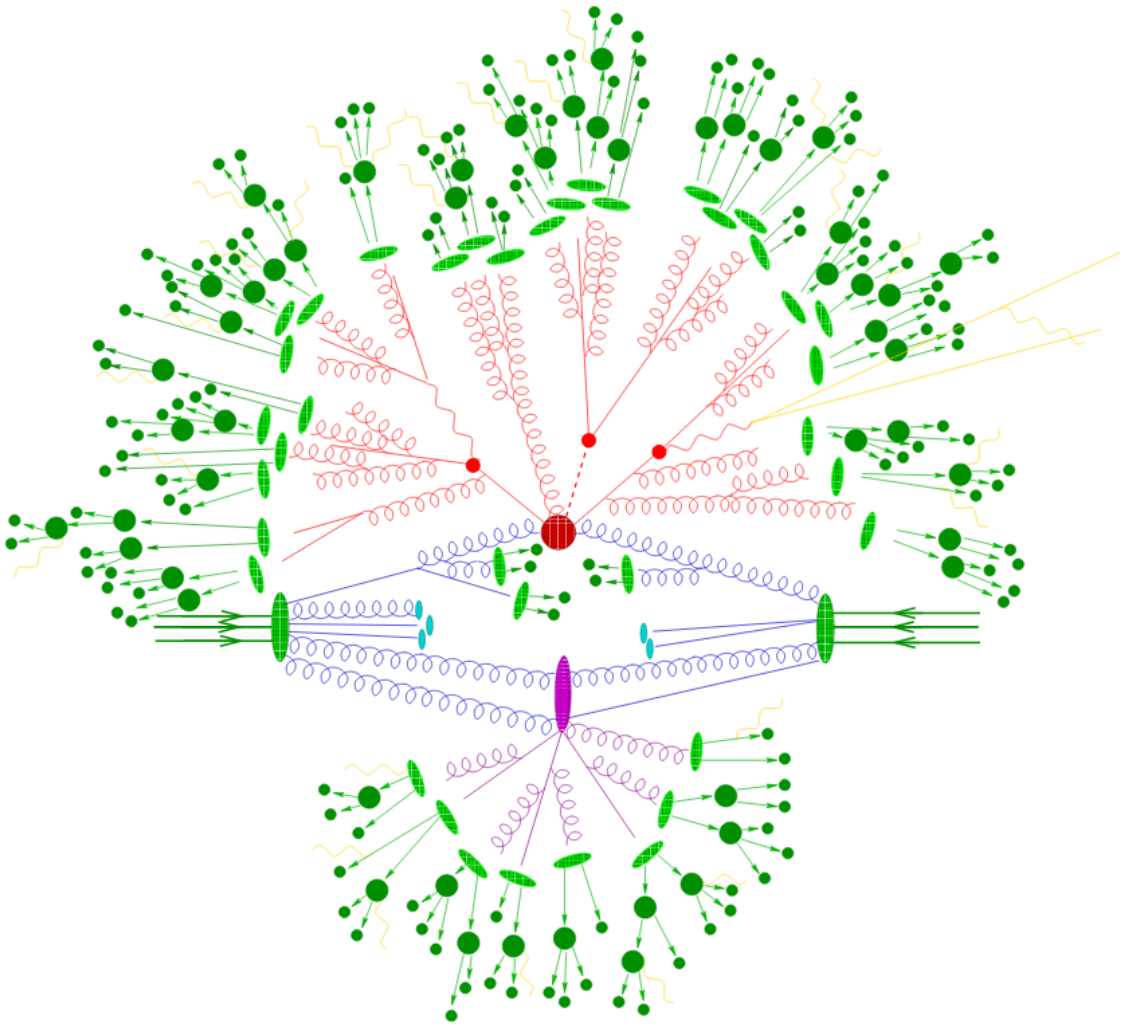


Figure 3.7: Hadron-hadron collision in a particle accelerator simulated by a Monte Carlo event generator. The large red dot at the center denotes the hard collision which is surrounded by a tree-like, also red structure representing the parton showers created by bremsstrahlung effects. The violet ellipse below shows a second hard collision of two partons. The light green blobs mark the transition of the colored parton state into the color neutral hadronic final state, which is followed by the decay processes of the hadrons (dark green bullets). Finally, the radiation of low energetic photons is depicted in yellow. Image from [54].

CHAPTER 4

TMVA & Boosted Decision Trees

TMVA (**T**oolkit for **M**ulti**V**ariate **A**nalysis) [55] offers a bunch of different multivariate classification and regression algorithms, integrated in the ROOT environment [56]. Section 4.1 gives a short overview of the training process using supervised learning [57] techniques, followed by the evaluation procedure in section 4.2. Afterwards, the MVA method of boosted decision trees, which is used for classification of signal and background events in the later part of this thesis, is introduced in section 4.3.

4.1 Training using the TMVA Factory

The first part of such a classification analysis is the training of the multivariate methods that should be used for discrimination. For that purpose, a TMVA `Factory` object is created which handles all necessary interaction between the user and the algorithms. It provides member functions for the selection of datasets for the training procedure, registration of training variables and booking of the MVA methods. First, the input events get split into a training and a test set, including the possibility of setting additional event weights. Since the treatment of events with negative weights as generated by some Monte Carlo generators can cause problems in the process, they are often ignored or the absolute value of the weight is used for training. By doing the event splitting in a randomized way, the test set allows for a cross check of the training performance after the training process which should be approximately equal on both datasets. If that is not the case, an effect of "overtraining" occurs where the used method (in the most extreme case) does not learn the characteristic features and properties of the different datasets (which are useful for classification), but basically memorizes the single events. This effect often appears when the number of degrees of freedom in the setup is too low, resulting in some methods being more susceptible for it than others. For boosted decision trees, the parameter which needs to be tuned in order to overcome overtraining usually is the number of nodes in the tree (see also section 4.3). Since with such a behavior the MVA method would not be applicable to a real analysis, this cross check is very important [55].
When all the inputs are specified, each method is trained, tested and evaluated before the factory creates method-specific weight files and a ROOT output file which contains the training results. Furthermore, it also provides a list of the input variables ranked by their discrimination power for the different methods [55]. An example workflow of the whole training procedure is graphically illustrated in the left part of figure 4.1.

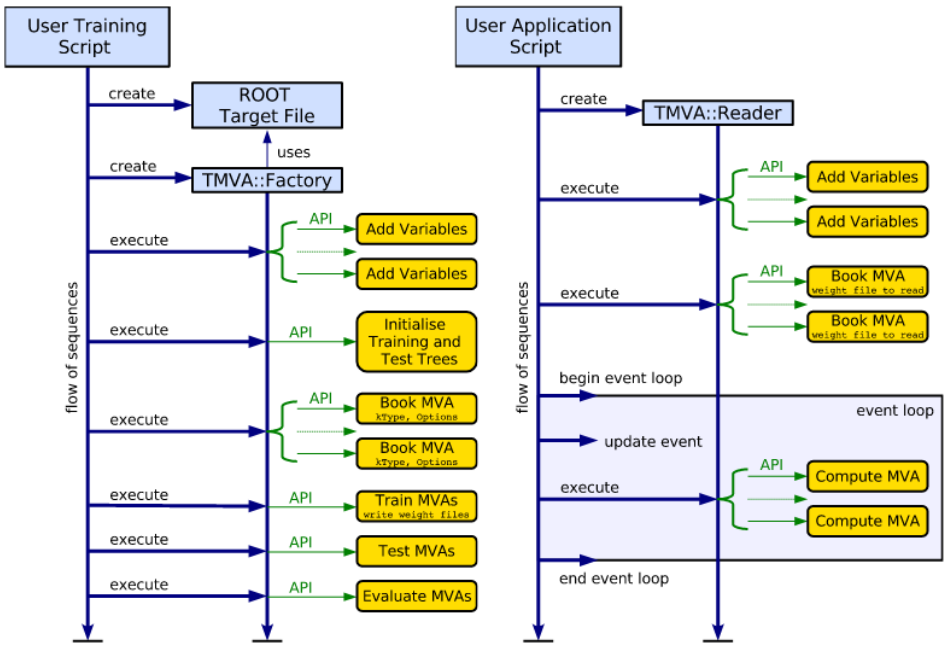


Figure 4.1: The left side shows an example for the training step using the TMVA `Factory`. Initialized and called by a user script (can be C++, Python, ROOT or similar), the factory object first reads the training variables (need to be in form of a `TFormula`), sets up the training and test ROOT trees and books the specified MVA methods with the given options. Then it trains, tests and evaluates all methods, creates the method-specific weight output files and stores further results in the ROOT target file which can be named by the user.
A sketch of the subsequent application step is depicted on the right side. In analogy to the factory object before, now a TMVA `Reader` object is created which evaluates the different MVA methods on the analysis samples and classifies their events as signal- or background-like. The list of variables added to the reader needs to match the one from the training step, but the list of MVA methods does not need to, as long as a weight file for the desired method exists. Afterwards, in the event loop each event is evaluated with respect to the booked methods and the classifier output distributions are filled. Image from [55].

4.2 Evaluation using the TMVA Reader

In the second step, which is performed independently from the first one, the training results are evaluated on the events of the analysis data samples (with unknown composition) by the TMVA `Reader` object. First, the variables from the training step and the desired MVA methods need to be registered. For this purpose the weight files produced at the end of the training step are used. Afterwards, for each method a classifier output variable can be created and is then filled by looping over all events in the analysis data samples [55]. The workflow of the evaluation step is shown in the right part of figure 4.1. In the generated classifier distributions, signal- and background-like events usually tend to peak at different positions which can be nicely seen in the example given in figure 4.2. Therefore, they can then be used as discriminating variables in the further analysis like other event variables (multiplicities, p_{T}, $E_{\mathrm{T}}^{\mathrm{miss}}$, ...) to optimize e.g. the signal-to-background ratio in a certain event channel or to build phase space regions which are enriched with either signal or background events, as it will be done in chapter 7 of this thesis.

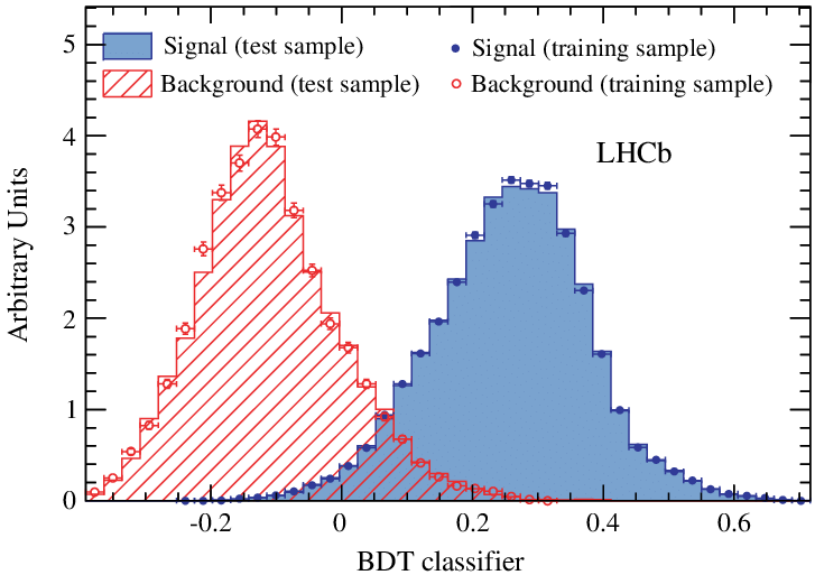


Figure 4.2: Example for a BDT classifier distribution from the LHCb experiment at CERN. Background events peak at a classifier value slightly smaller than −0.1 while signal-like events show a peak at approximately 0.3, resulting in a good separation between both samples. The distribution of the test events for both samples is superimposed on the training distribution and shows good agreement as well as hardly any indication of overtraining. Image from [58].

4.3 Decision Trees

A decision tree [55][59] is, compared to other machine learning techniques like neural networks [60][61], a relatively simple algorithm which can be used in particle physics for different classification tasks. Section 4.3.1 shortly describes the working principle of a single decision tree, while section 4.3.2 explains the expansion to a whole forest of decision trees, called boosting, as well as the two most common boosting algorithms *AdaBoost* and *GradientBoost*.

4.3.1 Basics

The basic principle of a decision tree is, as it might be inferred from its name, a repeated yes/no decision taken on a single variable at each node, leading to a tree-like structure with several "leaves" representing phase space hypercubes containing in majority either signal- or background-like events, as depicted in figure 4.3. Starting from the root node, for each yes/no decision all training variables are scanned with a certain step width over their whole range to find the variable x_i and cut value c_i which gives the best separation power between signal and background at the current node. This procedure is iterated until the maximum tree depth or minimum number of events in a leaf specified by the user is reached. In order to allow for a more useful scanning procedure, the input variables usually are transformed to a Gaussian distribution before [55].
There exist different splitting criteria (see [55], section 8.13.3 for more information) but since they all lead to a similar performance, only the default criterion used by TMVA, called *Gini Index*, is introduced here. It is defined by $p \cdot (1 - p)$ where p denotes the purity of the subset of events at the

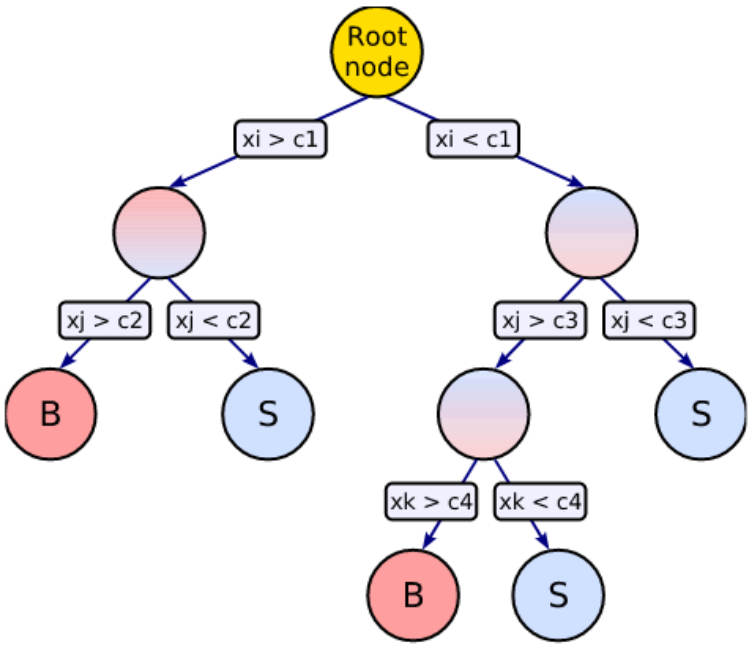


Figure 4.3: Sketch of a simple decision tree. Starting from the root node, the data sample gets split into signal- (S) and background-like (B) subsets until a stopping criterion is fulfilled. At each node, the variable x_i and cut value c_i yielding the best separation between S and B is chosen. Image from [55].

given node, i.e. $\frac{N_{\text{signal}}}{N_{\text{total}}}$. This leads to a criterion maximum at $p = 0.5$ which represents a subset with an equal number of signal and background events. Because a cut separating a pure background subset is as desirable as one yielding only signal events, the criterion is symmetric with respect to both event classes and falls off to zero for $p \to 0$ and $p \to 1$ [55].
Although this simple setup can already provide good results in classification problems, the tree-like structure has a major disadvantage, namely its instability to statistical fluctuations in the training sample. In case of a node where two variables yield similar separation power, a statistical fluctuation might result in the one variable to be chosen while without these fluctuation the other one would have been used for the split decision. The closer to the root node such a fluctuation occurs, the larger the impact on the decision tree since the whole following node structure including the used variables and cut values can be different in both cases [55].

4.3.2 Boosting

In order to minimize the effect of these statistical fluctuations, the single decision tree can be expanded to a whole forest of trees, typically a few hundred. An event is then classified as signal or background based on the majority of votes of the forest trees. All decision trees are trained successively on the same data samples where after the training of each tree a boosting is applied which basically modifies the weight of certain events in the samples to compensate for the statistical fluctuations. Hence, the final result then is a weighted mean of the decision trees in the forest. This procedure can significantly improve the separation performance compared to a single decision tree, but on the other hand the

possibility of an easy schematic interpretation of the tree gets lost [55]. The two most common boosting algorithms for decision trees which are also used in this analysis are shortly explained below.

Adaptive Boost

The adaptive boost (*AdaBoost*) is the most frequently used boosting algorithm for classification problems. The first tree in the forest is trained using the nominal event weights for each sample but for all remaining trees the weights are modified. For each subsequent tree, all events misclassified in the previous tree get an additional boost weight α which is calculated from the rate err of wrong classified events in the tree before as follows [55]:

$$\alpha = \frac{1 - \mathrm{err}}{\mathrm{err}} \quad . \tag{4.1}$$

Afterwards, a renormalization is applied to the remaining events to keep the total sum of weights in each sample constant. Defining the output of an individual classifier as $h(x) = +1$ for signal and -1 for background, respectively, with x being the tuple of input variables, the boosted event classification $y_{\mathrm{Boost}}(x)$ can be written as [55]:

$$y_{\mathrm{Boost}}(x) = \frac{1}{N_{\mathrm{collection}}} \cdot \sum_{i}^{N_{\mathrm{collection}}} \ln\left(\alpha_i\right) \cdot h_i(x) \quad . \tag{4.2}$$

In this equation, it is summed over all classifiers in the collection. Small values of $y_{\mathrm{Boost}}(x)$ indicate a background-like event while signal events tend to have a large $y_{\mathrm{Boost}}(x)$. The loss function L representing the deviation between the model response $F(x)$ and the true value y determined from the training sample (see [62] for a general short introduction) is here given by

$$L(F(x), y) = e^{-F(x)y} \tag{4.3}$$

where $F(x)$ is assumed to be a weighted sum of parametrized base functions $f(x; a_m)$ which are called the "weak learners" [55]:

$$F(x; P) = \sum_{m=0}^{M} \beta_m f(x; a_m); \quad P \in \{\beta_m; a_m\}_0^M \quad . \tag{4.4}$$

The boosting is then carried out by minimizing the loss function L. Doing so, *AdaBoost* yields a very good performance on weak classifiers, i.e. single decision trees with a depth of not larger than 2-3 which makes it a good choice for the setup of this analysis.

Gradient Boost

The shortcoming of an exponential loss function as used for the *AdaBoost* is its instability with respect to mislabelled data points or outliers, leading to a decreased performance for setups with large noise. The *GradientBoost* algorithm therefore aims to tackle this issue with the choice of the slightly more

robust binomial log-likelihood loss function instead [55]:

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

For this loss, the minimization of the boosting algorithm needs to be done using a steepest-descent approach. For that purpose first the current gradient of the loss function is calculated, thereafter a regression tree is built whose leaf values are tuned in a way that they match the mean value of the gradient in the different regions defined by the structure of the tree. The set of decision trees minimizing the loss function is then obtained by iteration of this process [55].

CHAPTER 5

Multi-Tau Event Identification

As already stated in chapter 3, the reconstruction and identification of hadronically decaying τ-leptons is of special importance for the investigation of a 4τ final state as performed in this analysis. Consequently, dedicated methods are required to separate true tau events from those faked by QCD multijet or other background processes (referred to as fake taus). Section 5.1 describes the "classic" identification approach used in ATLAS, while section 5.2 shortly introduces a new identification method for multi-tau events developed in [8] which is based on multiplied probabilities. All text explanations given in that section originate from this reference. Moreover, some plots from [8] have been reproduced for this chapter, but they might look slightly different due to some settings not being exactly equal and should only serve as qualitative illustration. Further detailed information concerning the development procedure and expected event yields can also be found in [8].

5.1 Tau Lepton Identification in ATLAS

Reconstruction

Due to its short lifetime of $\approx 10^{-13}$ s [27], the τ-lepton decays rather quickly and needs to be reconstructed from its visible decay products (referred to as $\tau_{\text{had-vis}}$) in the detector. Furthermore, the decay length of 87 µm [63] leads to a displacement of the tau decay vertex (TV) compared to the primary collision point. Jets reconstructed by the anti-k_t algorithm [43] with a distance parameter of $R = 0.4$ are used as seeds for the $\tau_{\text{had-vis}}$ reconstruction in combination with calorimeter cluster cells (as described in section 3.3.1), if they fulfill $p_{\text{T}} > 10$ GeV and $|\eta| < 2.5$. Then, the TV is identified from an algorithm evaluating all tau candidate tracks in the region $\Delta R < 0.2$ around the jet seed direction and summing up their transverse track momenta. The vertex with the largest matched p_{T} fraction is chosen to be the TV. Track candidates further need to satisfy $p_{\text{T}} > 1$ GeV, at least two hits in the pixel detector and a minimum of seven hits in the pixel detector and the subsequent SCT combined. The reconstruction efficiency for 1-prong and 3-prong tau leptons depending on their truth visible transverse momentum is depicted in figure 5.1. This plot only considers tau candidates which fulfill the final requirements of $p_{\text{T}} > 20$ GeV, $|\eta| < 2.5$ and η outside of the transition region of $1.37 < |\eta| < 1.52$ between the inner detector and the electromagnetic calorimeter [64].

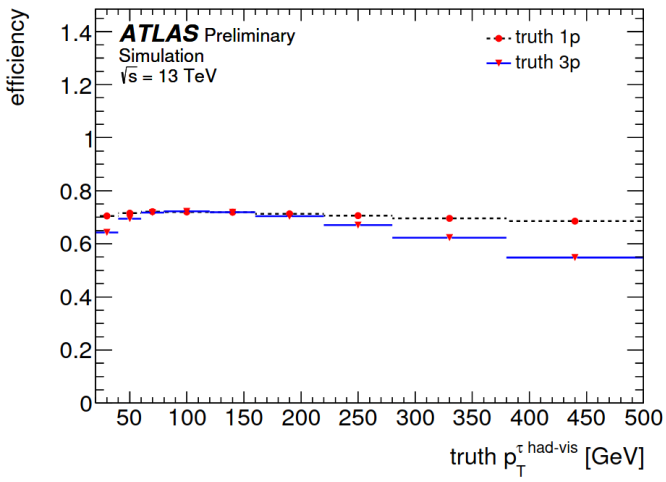


Figure 5.1: Efficiency for reconstruction of all charged tracks stemming from the tau decay for 1-prong and 3-prong taus as a function of truth $\tau_{\text{had-vis}}$ p_{T}. Image from [64].

Identification

Since reconstruction of tau candidates on its own does not provide a significant rejection against hadronic background, this has to be achieved in an additional step. ATLAS uses boosted decision trees or recurrent neural networks [65] for that purpose. In this analysis, the BDT approach is chosen where the single trees are trained on $Z/\gamma* \rightarrow \tau\tau$ (signal) and di-jet (background) events for 1-prong and 3-prong decays separately. The discriminating variables used for the training process are based on the reconstructed tracks as well as on the cluster information from the calorimeters and are further corrected to account for pile-up effects. They (and their description) are directly taken from [64], can be also found in [8] and are listed below:

- **Central energy fraction (f_{cent}):** Fraction of the calorimeter transverse energy deposited in the region $\Delta R < 0.1$ with respect to all energy deposited in the region $\Delta R < 0.2$ around the $\tau_{\text{had-vis}}$ candidate. It is calculated by summing the energy deposited in all cells belonging to TopoClusters with a barycenter in these regions, calibrated at the electromagnetic energy scale.

- **Leading track momentum fraction ($f^{-1}_{\text{leadtrack}}$):** The transverse energy sum, calibrated at the electromagnetic energy scale, deposited in all cells belonging to TopoClusters in the core region of the $\tau_{\text{had-vis}}$ candidate, divided by the transverse momentum of the highest-p_{T} charged particle in the core region.

- **Track radius ($R^{0.2}_{\text{track}}$):** p_{T}-weighted ΔR distance of the associated tracks to the $\tau_{\text{had-vis}}$ direction, using only tracks in the core region.

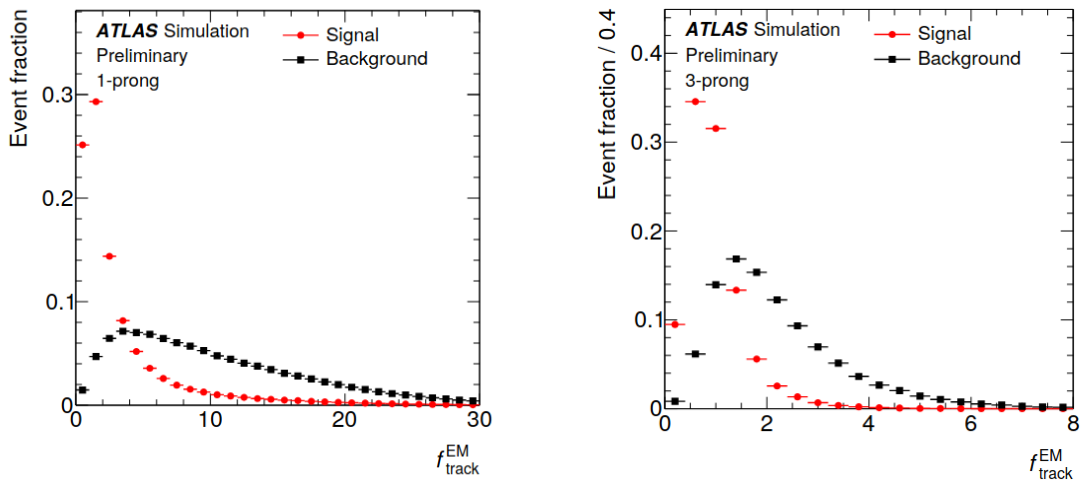
- **Leading track IP significance ($|S_{\text{leadtrack}}|$, 1-prong only):** Absolute value of the transverse impact parameter of the highest-p_{T} track in the core region, calculated with respect to the TV, divided by its estimated uncertainty.

- **Fraction of track-p_{T} in the isolation region ($f^{\text{track}}_{\text{iso}}$, 1-prong only):** Scalar sum of the p_{T} of

tracks associated with the $\tau_{\text{had-vis}}$ candidate in the region $0.2 < \Delta R < 0.4$ divided by the sum of the p_{T} of all tracks associated with the $\tau_{\text{had-vis}}$ candidate.

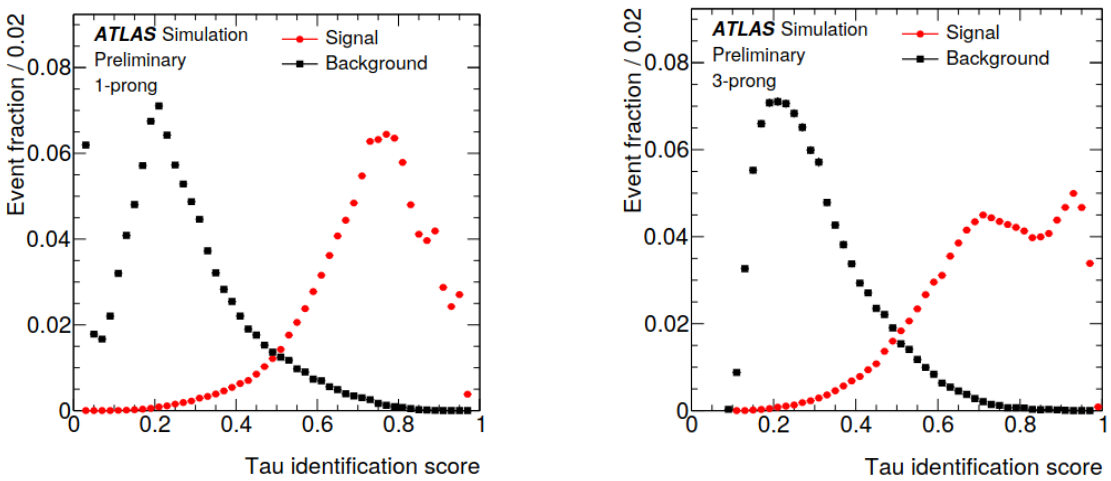
- **Maximum ΔR (ΔR_{Max}, 3-prong only):** The maximum ΔR between a track associated with the $\tau_{\text{had-vis}}$ candidate and the $\tau_{\text{had-vis}}$ direction. Only tracks in the core region are considered.
- **Transverse flight path significance ($S_{\text{T}}^{\text{flight}}$, 3-prong only):** The decay length of the secondary vertex (vertex reconstructed from the tracks associated with the core region of the $\tau_{\text{had-vis}}$ candidate) in the transverse plane, calculated with respect to the TV, divided by its estimated uncertainty.
- **Track mass (m_{track}, 3-prong only):** Invariant mass calculated from the sum of the four-momentum of all tracks in the core and isolation regions, assuming a pion mass for each track.
- **Fraction of EM energy from charged pions ($f_{\text{EM}}^{\text{track-HAD}}$):** Fraction of the electromagnetic energy of tracks associated with the $\tau_{\text{had-vis}}$ candidate in the core region. The numerator is defined as difference between the sum of the momentum tracks in the core region and the sum of cluster energy deposited in the hadronic part of each TopoCluster associated with the $\tau_{\text{had-vis}}$ candidate. The denominator is the sum of the cluster energy deposited in the electromagnetic part of each TopoCluster associated with the $\tau_{\text{had-vis}}$ candidate.
- **Ratio of EM energy to track momentum ($f_{\text{track}}^{\text{EM}}$):** Ratio of the sum of cluster energy deposited in the electromagnetic part of each TopoCluster associated with the $\tau_{\text{had-vis}}$ candidate to the sum of the momentum of tracks in the core region.
- **Track-plus-EM-system mass ($m_{\text{EM+track}}$):** Invariant mass of the system composed of the tracks and up to two most energetic EM clusters in the core region, where EM cluster energy is the part of TopoCluster energy deposited in the presampler and first two layers of the EM calorimeter. The four-momentum of an EM cluster is calculated assuming zero mass and using TopoCluster seed direction.
- **Ratio of track-plus-EM-system to p_{T} ($p_{T}^{\text{EM+track}}/p_{\text{T}}$):** Ratio of the $\tau_{\text{had-vis}}$ p_{T}, estimated using the vector sum of track momenta and up to two most energetic EM clusters in the core region to the calorimeter-only measurement of $\tau_{\text{had-vis}}$ p_{T}.

Figure 5.2 exemplary shows the $f_{\text{track}}^{\text{EM}}$ variable for 1-prong and 3-prong taus, while the resulting tau identification scores are depicted in figure 5.3. The identification efficiency can then be defined as the fraction of 1-prong (3-prong) hadronic tau decays which are reconstructed as 1-track (3-track) $\tau_{\text{had-vis}}$ candidates and also pass the BDT selection criteria. Multiplying this value with the p_{T}-dependent reconstruction efficiency from the previous section finally yields the combined tau signal efficiency. In order to make this signal efficiency independent from p_{T} again, the cut values on the tau identification scores need to be chosen as a function of p_{T} as well. Therefore, it is not possible to provide a fixed cut value on the score distribution to always obtain the same signal efficiency. This problem is usually solved by a transformation of the score in a way that it directly represents the identification efficiency of $\tau_{\text{had-vis}}$ candidates. The transformed score $\tau_{\text{trans}}^{\text{jetBDT}}$ is running from 0 (jet-like) to +1 (tau-like). A value of e.g. $\tau_{\text{trans}}^{\text{jetBDT}} = 0.4$ means that 40 % of the reconstructed taus do not exceed this value, resulting in a signal efficiency of 60 % [64].



(a) 1-prong (b) 3-prong

Figure 5.2: Distribution of the $f_{\text{track}}^{\text{EM}}$ BDT training variable used for tau identification, split in 1-prong and 3-prong decays. Tau events are shown in red, while the black curves represent jet events. Image from [64].



(a) 1-prong (b) 3-prong

Figure 5.3: Tau identification score obtained from the BDT for tau (signal) and jet (background) events for 1-prong (left) and 3-prong (right) decays, respectively. Image from [64].

Tau identification in ATLAS uses three different working points, labeled *loose*, *medium* and *tight*, corresponding to a certain identification efficiency (and transformed BDT score) each. They are summarized in table 5.1 together with the additional, rarely used *very loose* working point [64]. While choosing one of them results in good signal efficiencies for many ATLAS analyses (which mostly do not investigate final states with more than two taus), such a fixed working point does not yield a sufficient performance for the identification of a four tau final state as considered in this thesis. The low number of expected events for such a diboson decay channel would be further decreased by the

requirement that each of the four τ-leptons would have to pass the BDT score corresponding to the selected working point [8]. Thus, a new tau identification method for multi-tau events in the $hh \to 4\tau$ channel has been developed in [8] which is briefly described in the next section and will be used in this analysis as well.

Working Point	Signal efficiency		Background rejection	
	1-prong	3-prong	1-prong	3-prong
Tight	60 %	45 %	40	400
Medium	75 %	60 %	20	150
Loose	85 %	75 %	12	61
Very loose	95 %	95 %	5.3	11.2

Table 5.1: Identification efficiencies of $\tau_{\text{had-vis}}$ candidates and corresponding background rejection achieved with the BDT for the different defined tau working points. Taken from [65].

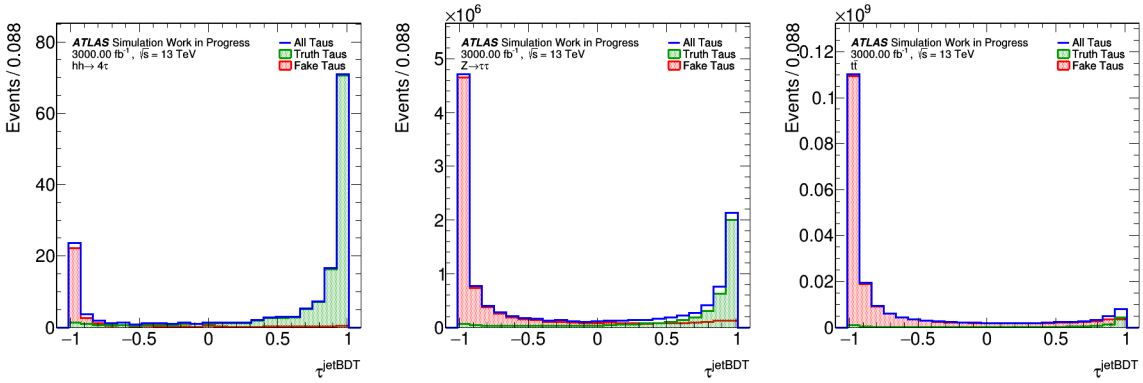
5.2 Multiplied Probability Method

The main idea of a better identification algorithm of multi-tau events is to not require the single taus to fulfill a certain working point anymore. Instead, the decision whether an event is identified as a four-tau event is made based on the combined τ^{jetBDT} identification score information from the τ-leptons provided in the previous section, which can be motivated from figure 5.4 quite well. It shows the τ^{jetBDT} score distribution of the $hh \to 4\tau$ signal and two possible background processes, split into truth and fake taus. While there is a considerable amount of truth taus for the signal, the backgrounds have only a moderate to vanishing contribution of such events and are dominated by fake taus. Nevertheless, there is also a non-negligible fraction of fake tau events for the signal process. The requirement of a fixed working point for all four τ-leptons would now lead to two possible problems:

- For the signal process, there is a high chance for at least the fourth tau to not exceed the necessary τ^{jetBDT} score of a medium or even loose working point. Consequently, a significant fraction of $hh \to 4\tau$ events might be lost, although three of four taus are reconstructed very well and have a relatively high score.

- For the background, there is still a considerable number of events in the region of medium τ^{jetBDT} scores, i.e. $|\tau^{\text{jetBDT}}| \lesssim 0.7$, since the absolute number of background events exceeds those of signal events by far. Hence, such events with four medium score taus would be accepted as long as every single tau exceeds the cut value given by the chosen working point, even if its τ^{jetBDT} score is really close to the threshold.

In order to overcome these problems, a score combining the information from the single taus needs to have two important properties. First, it should not reject events in which the majority of taus is reconstructed very well and only one (or maybe two) τ-leptons have a medium or even low τ^{jetBDT} score. On the other hand, it needs to provide rejection power against events which are dominated by medium score taus and thus are more background-like. For that purpose, a new quantity τ^{prob} is



(a) $hh \to 4\tau$ signal process (b) $Z \to \tau\tau$ background process (c) $t\bar{t}$ background process

Figure 5.4: τ^{jetBDT} score distribution of different physics processes for an integrated luminosity of 3 000 fb, split into truth and fake taus. For the signal process (a) the fraction of truth tau events is approximately 75 %. Background processes producing two truth taus as $Z \to \tau\tau$ (b) show a moderate amount of truth taus, while in a process like $t\bar{t}$ production (c) almost all taus are faked by jets.

defined in the following way:

$$\tau^{\text{prob}} = \frac{N^{\text{truth}}\left(\tau^{\text{jetBDT}}\right)}{N^{\text{truth}}\left(\tau^{\text{jetBDT}}\right) + N^{\text{fake}}\left(\tau^{\text{jetBDT}}\right)} \quad . \tag{5.1}$$

Given that N^{truth} and N^{fake} are the numbers of truth and fake taus for a certain τ^{jetBDT} value, respectively, τ^{prob} thus can be interpreted as the probability of a tau candidate to be a truth tau. Figure 5.5 shows the distribution of τ^{prob} as a function of τ^{jetBDT} for the signal and different background processes. As equation 5.1 directly depends on the numbers of truth and fake taus in the single τ^{jetBDT} bins, one can expect the distributions to differ significantly from each other, based on the amount of truth taus which usually emerge from the corresponding process. The plot confirms this expectation.
Furthermore, the impact of the sorting order of the four tau leptons with respect to several quantities (p_{T}, η, τ^{jetBDT}, 1- or 3-prong) on the τ^{prob} distribution has been studied in broad detail in [8]. For each sorting algorithm, the taus are separated by the application of cuts which split the events into different phase space regions in the sorting variable, e.g. bins in p_{T}. The highest significance gain $S/\sqrt{S+B}$ with S being the number of signal and B the number of background events could be achieved with a combination of the τ^{prob} variable and the prong-based sorting method [8]. This result can be explained by the fact the τ^{jetBDT} scores introduced in section 5.1 are already defined separately for 1-prong and 3-prong taus, while they are relatively flat with respect to most other observables. A better performance of the prong-based sorting method compared to the other ones might therefore be somehow expected [8].
The new variable used for discrimination of truth-tau events against τ-leptons faked by jets is now constructed as the product $\prod \tau_i^{\text{prob}}$ of the single tau probability scores τ^{prob} defined in equation (5.1). Figure 5.6 shows its separation power on a Sherpa $\ell\ell\ell\ell$ event sample (cf. tab. A.6). It can be nicely seen that the base-10 logarithm of the designed probability product yields a large separation between events with many and those with less truth taus. Although the discrimination performance of truth

from fake tau events, compared to a cut on every single τ^{jetBDT} score, does not increase significantly for the $hh \rightarrow 4\tau$ signal sample (cf. fig. 5.7(a)), a large gain can be observed for background events, as exemplary shown for the $t\bar{t}$ process in figure 5.7(b). Furthermore, the method of multiplied probabilities also exceeds the "classic" τ^{jetBDT} score cut-based approach in terms of significance. For a signal loss of 61 % (corresponding to the requirement of four taus fulfilling the *loose* working point), the expected significance can be doubled [8]. Hence, cutting on the tau probability product seems to be very suitable for the separation of signal from background events in multi-tau final states. Consequently, it will also be used in the analysis performed in this thesis.

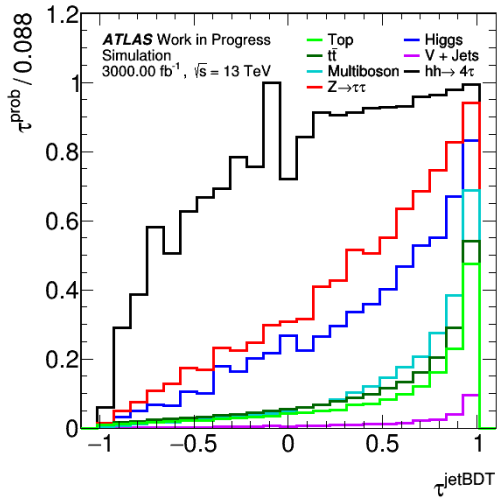


Figure 5.5: The probability value τ^{prob} as function of τ^{jetBDT} shows large differences in the distribution for the different physics processes.

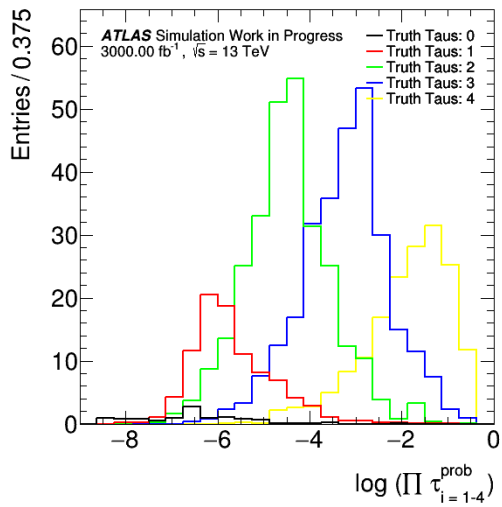
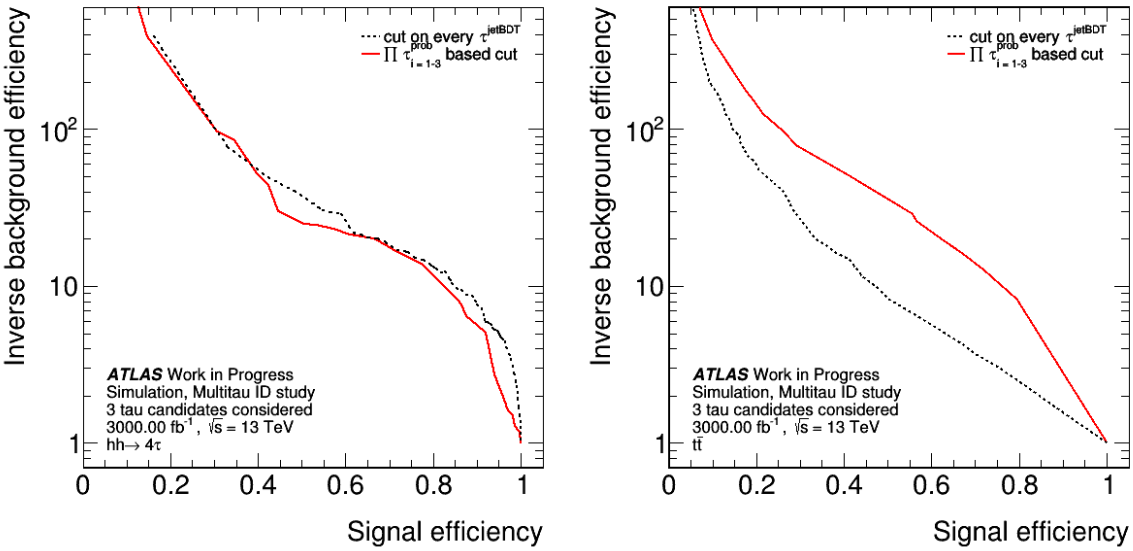


Figure 5.6: Qualitative separation power of the tau probability product for multi-tau events with different numbers of truth taus, exemplary shown for a Sherpa $\ell\ell\ell\ell$ sample. Events with a high number of truth taus peak at values close to zero, while fake-enriched events accumulate at larger negative values.



(a) $hh \rightarrow 4\tau$ signal process

(b) $t\bar{t}$ background process

Figure 5.7: The plots show the rejection power against fake-tau events as a function of truth-tau signal efficiency (ROC curve) for different physics processes and tau identification methods. The "classic" approach, corresponding to a cut on the τ^{jetBDT} score of every single tau candidate, is shown as a dashed black line, while the solid red curve shows the performance for the method of multiplied probabilities. While no significant difference can be observed for the signal process, the probability-based cut clearly outperforms the "classic" approach when being applied on the $t\bar{t}$ background process.

CHAPTER 6

Data Samples & Background Estimation

Given the identification method for multi-tau events presented in chapter 5, the following sections will now focus on the general analysis setup. The analysis is performed on LHC run 2 2017 data with an integrated luminosity of 46.9 fb^{-1}, recorded with the ATLAS detector at a center-of-mass energy of 13 TeV. At first, section 6.1 gives an overview of observables used later on in the analysis, before in section 6.2 the signal samples for the $hh \rightarrow 4\tau$ and $ZZ \rightarrow 4\tau$ processes, which are modeled by Monte Carlo simulations, are briefly introduced. Afterwards, the relevant backgrounds for a 4τ final state are discussed in section 6.3, followed by the applied preselection requirements and resulting data modeling in section 6.4. Finally, the estimation of background events with τ-leptons exclusively faked by jets and its further improvement are treated in the sections 6.5 and 6.6.

6.1 Observable Definitions

In order to allow for a better understanding of the plots shown in the next sections and chapters, first the most frequently used and not self-explaining observables are explained below:

- $\boldsymbol{m_{\tau\tau}}$**:** Mass of the ditau system formed by the respective τ-leptons. The tau candidates are sorted with respect to their τ^{jetBDT} score, i.e. τ_1 is always the tau with the highest τ^{jetBDT} score.
- $\boldsymbol{\Delta R_{\tau\tau}}$**:** Three-dimensional distance measure between the given τ-leptons.
- $\boldsymbol{N^{\tau}_{\text{tracks}}}$**:** Number of charged tracks in the tau decay products.
- $\boldsymbol{\log\left(\prod \tau^{\text{prob}}_{\text{i=1-3/4}}\right)}$**:** Probability product of three or four taus, respectively. If there is no index "i=1-3" or "i=1-4" added to the name of the observable or if the index is "i=1-3,4", the 3-tau product is chosen for events from the $3h1l$ channel while the 4-tau product is used for the remaining events from the $4h0l$ channel.
- $\boldsymbol{m_{\tau_1\tau_2\tau_3\tau_4}}$**:** Invariant mass of all four ($4h0l$) or three tau candidates and the light lepton ($3h1l$). They form two ditau systems which again build the di-Higgs (chap. 6) or di-Z (chap. 7 & 8) system.
- $\boldsymbol{m^{\text{low/high}}_{\tau\tau}}$**:** With four taus, there are two possibilities to combine one positively charged and one negatively charged tau each to a ditau system. Thus, one can differentiate between one

system where the taus are assigned correctly and one where the wrong taus are combined[1]. The observable denotes the mass of the ditau system with correctly assigned τ-leptons and the lower or higher mass of the two systems, respectively.

- $\boldsymbol{m_{\tau\tau}^{\text{low/high, false}}}$**:** Mass of the ditau system with falsely assigned τ-leptons and the lower or higher mass of the two systems, respectively.
- $\boldsymbol{\Delta R_{\tau\tau}^{\text{low/high}}}$**:** Distance measure between the two taus in the ditau system with correctly assigned τ-leptons and the lower or higher ΔR value, respectively.
- $\boldsymbol{\Delta R_{\tau\tau}^{\text{low/high, false}}}$**:** Distance measure between the two taus in the ditau system with falsely assigned τ-leptons and the lower or higher ΔR value, respectively.
- $\boldsymbol{\sum p_{\mathrm{T}}(\mathrm{lep}_{1-4}, 3h1l)}$**:** Scalar sum $|p_{\mathrm{T}}(\tau_1)|+|p_{\mathrm{T}}(\tau_2)|+|p_{\mathrm{T}}(\tau_3)|+|p_{\mathrm{T}}(\ell)|$ of the three tau candidates and the light lepton which form the two ditau systems in the $3h1l$ channel.

6.2 Signal Processes

In the Standard Model, a final state with four truth τ-leptons originating from not more than two particles can only be produced from diboson decays, either from two Higgs or Z bosons. Both processes have a very small cross section (cf. tab. A.1 & A.6) and suffer from a large background exceeding the expected signal event yields by many orders of magnitude. The $hh \to 4\tau$ process has already been used in the previous chapter to demonstrate the power of the method of multiplied probabilities for tau identification and will further serve as a representative signal for the background estimation techniques shown in the subsequent sections. Chapter 7 and 8 then focus on the discrimination of $ZZ \to 4\tau$ events against their background and a measurement of the cross section of this process.

The $hh \to 4\tau$ Process

As stated in table 2.4, 6.4 % of all produced Higgs bosons decay into a pair of τ-leptons. The probability for a di-Higgs decay into four taus is thus given by $(6.4\,\%)^2 = 0.41\,\%$. Although there are other final states with much higher branching ratios including b-quarks or vector bosons, like $hh \to b\bar{b}\tau\tau$ or $hh \to WW^*\tau\tau$, they all come with a significant larger hadronic background than the 4τ state. Hence, the $hh \to 4\tau$ channel might be a suitable choice despite its small cross section and the identification method for multi-tau events introduced in section 5.2 is a key to probe the potential of this channel.
According to the possibility of a leptonic or hadronic decay for each of the four τ-leptons (cf. fig. 2.10), the $hh \to 4\tau$ channel can be further split into different sub-channels which are depicted in figure 6.1. In the most frequent case with a probability of 38 %, three taus decay hadronically while the fourth one decays into a light lepton and the additional neutrinos. Consequently, this channel is labeled $3h1l$. It is followed by the $2h2l$ (31 %) and the $4h0l$ (18 %) channels. The two remaining final states only occur with a probability of at total 13 %. In order to make the best possible use of the tau

[1] Of course a differentiation between the correct and false assigned combinations is only possible using truth-level information. On reconstruction level, the pairing is done based on the minimization of the distance ΔR between the possible tau pairs (cf. [8], sec. 6.3). The "false" combination in the following thus refers to the case where the tau candidates are paired opposite to the combination resulting from the minimum distance approach.

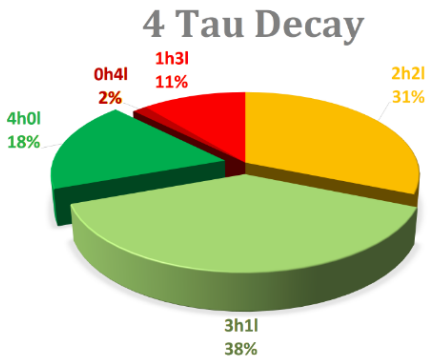


Figure 6.1: Possible sub-channels for the $hh \rightarrow 4\tau$ process and their branching ratios. They are labeled according to their number of hadronically (h) and leptonically (l) decaying taus. Image from [8].

identification method of multiplied probabilities, only the $4h0l$ and $3h1l$ channels will be investigated in this analysis[2]. The used signal sample has been generated with MC@NLO [52] while the particle showering was simulated with Herwig [66]. Further details can be found in table A.1 in appendix A.

The $ZZ \rightarrow 4\tau$ Process

In comparison to the Higgs boson, the branching ratio of the Z into two τ-leptons is only 3.4 % (cf. tab. 2.3), resulting in a probability of 0.12 % for a ZZ decay into four taus. Nevertheless, the tau branching ratios of course remain unchanged, so that figure 6.1 applies to the $ZZ \rightarrow 4\tau$ channel as well. Thus, the analysis towards the measurement of the cross section of this process in chapter 7 and 8 will also be restricted to the $4h0l$ and $3h1l$ sub-channels.
Since there was no exclusive $ZZ \rightarrow 4\tau$ data sample available with the start of the analysis, it had to be constructed separately. For that purpose, a $\ell\ell\ell\ell$ sample generated with the Sherpa generator [50] is used which contains Standard Model processes with four leptons in the final state. The $hh \rightarrow 4\ell$ propagator has been deactivated for the simulation here in order to ensure that the ZZ decay is the only process in the sample resulting in a final state with four correlated truth τ-leptons. Then, a $ZZ \rightarrow 4\tau$ sample can be easily cut out from this $\ell\ell\ell\ell$ sample by requiring at least three or four truth hadronic tau candidates for the $3h1l$ and $4h0l$ channel, respectively. All remaining events which do not fulfill this requirement are grouped into a new $\ell\ell\ell\ell$ sample which contributes to the background processes described in section 6.3.

6.3 Monte Carlo Background Processes

The outstanding characteristic of the final state chosen for this analysis is the production of four τ-leptons. As explained before, the Standard Model background processes in the 4τ channel mostly come up with at least one tau faked by a jet or light lepton. In the most signal-unlike case, all four taus are produced by QCD interactions and can be completely uncorrelated. The estimation of this special QCD multijet background, which usually can not be taken from Monte Carlo simulations but has to

[2] The $2h2l$ channel is also further contaminated by $hh \rightarrow WW^*\tau\tau$ events whose cross section is a lot larger than the one of the $hh \rightarrow 4\tau$ process.

be modeled directly from the recorded data, will be discussed separately in section 6.5. The main other physical processes contributing to the expected background are briefly explained below, detailed information concerning the used data samples can be found appendix A:

- **V + Jets** includes different processes of single vector boson production, i.e. W or Z bosons which then decay into a charged lepton and a neutrino or two charged light leptons, respectively. In that case, the leptons are misidentified as taus and additional jets fake the existence of two more τ-leptons.
- In all **Higgs** background processes a single Higgs boson decays into $\tau\tau$, where further taus either stem from additional vector boson and top quark decays or are faked by light leptons and QCD jets.
- Another background is coming from **$Z \rightarrow \tau\tau$ decays**, where again the two remaining taus are faked by jets.
- Production of multiple vector bosons, referred to as **Multiboson** from now on, decaying into taus, falsely reconstructed light leptons or jets is an important background as well. Di- and triboson processes decaying into different combinations of leptons and jets contribute to this background.
- The **$t\bar{t}$ background** includes decays of two top quarks with additional jets or light leptons misidentified as τ-leptons again.
- The same final state can be produced by **single top** decays together with light leptons or the decay of an additional vector boson.

In order to estimate the impact and expected event yields of these different processes, a truth-level study has been carried out in [8] for an integrated luminosity of $3\,000\,\mathrm{fb}^{-1}$. After application of a few simple cuts (cf. tab. 4.1 there), the dominant background contribution for the 4τ final state is expected to come from Higgs and multiboson processes and a significance of $S/\sqrt{S+B} = 1.94$ for the $hh \rightarrow 4\tau$ channel can be reached. However, the picture looks different on reconstruction level since the underlying truth information from the Monte Carlo simulations can not be exploited there. Thus, the next section will focus on how to choose some suitable preselection cuts which on the one hand reduce the expected background but do not reject too many of the rare signal events.

6.4 Preselection

The starting point of this analysis in terms of datasets are the so-called `DAOD`[3] files on which already a few cuts are placed. These cuts for the used derivations `HIGG4D2` and `HIGG4D3` are summarized in table 6.1. While for `HIGG4D2` at least one tau candidate and light lepton is demanded, `HIGG4D3` requires two tau candidates and additionally has a light lepton veto. Thus, these derivations basically

[3] Experimental data recorded with the ATLAS detector is first stored in *Analysis object data* (`AOD`) type files. Since their size usually is about a few petabytes, they are shrunk for the analysis of a certain physics process. The result of this procedure are *Derived AODs* (`DAOD`) files with a size in the order of terabyte which in a next step are again slimmed to a few gigabyte. This last step is carried out with the `xTauFramework` [67] which produces so-called n-tuples in form of a ROOT tree [8].

represent the $3h1l$ and $4h0l$ sub-channels of the analysis, respectively. Due to the very small expected signal yields compared to the overall background, further preselection cuts are kept as loose as possible. In the step of n-tuple production, all tau candidates are (in addition to the DAOD cuts) required to have a transverse momentum of at least 20 GeV, either one or three charged tracks in the tau core region and must lie in the accessible η-region of the detector. On analysis level, finally exactly four charged leptons are demanded from which two must have electric charge +1 and two −1. According to the $3h1l$ and $4h0l$ analysis sub-channels, which are treated together, the number of light leptons is restricted to be ≤ 1. The combination of these two cuts automatically ensures the existence of 3 or 4 tau candidates in the respective sub-channels. These cuts are summarized in table 6.2 and complete the applied preselection requirements.

Derivation	HIGG4D2	HIGG4D3
Object related	$N_\tau \geq 1$, $N_e + N_\mu \geq 1$	$N_\tau \geq 2$, $N_e + N_\mu = 0$
Momentum related (cut values in GeV)	$p_{\mathrm{T}}^{\tau} \geq 23$ and ($p_{\mathrm{T}}^{\mu} \geq 12$ or $p_{\mathrm{T}}^{e} \geq 15$) or $p_{\mathrm{T}}^{\tau} \geq 18$ and ($p_{\mathrm{T}}^{\mu} \geq 18$ or $p_{\mathrm{T}}^{e} \geq 22$)	$p_{\mathrm{T}}^{\tau_1} \geq 33$ and $p_{\mathrm{T}}^{\tau_2} \geq 23$
Object identification related	e = Medium or μ = Good	τ_1 = Loose or τ_2 = Loose
Tau related	$N_{\mathrm{tracks}}^{\tau} = 1$ or 3 , $\lvert q_\tau \rvert = 1$	$N_{\mathrm{tracks}}^{\tau_{1,2}} \in [1, 8]$, $\lvert q_{\tau_{1,2}} \rvert = 1$

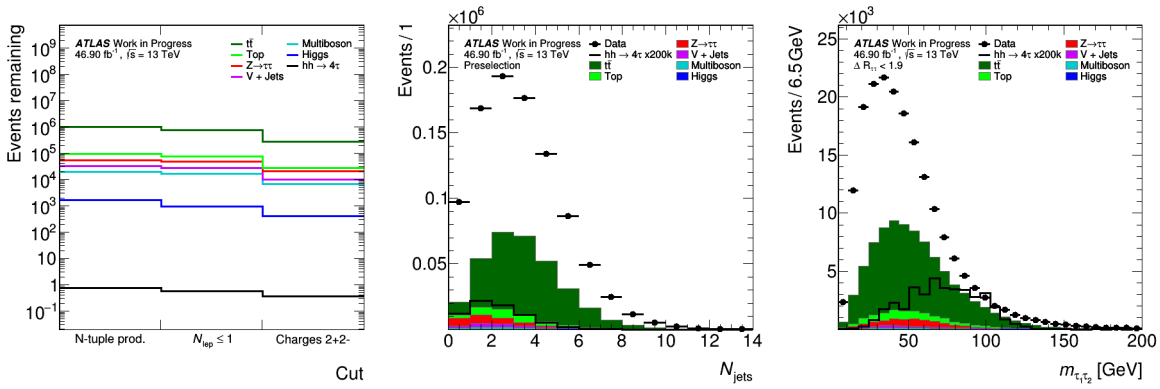
Table 6.1: The two different DAOD types used in this analysis and the requirements on a event to be accepted for the certain derivation sample are shown. Taken from [8], tab. 4.2.

Application step	Cut criteria
N-tuple production	$p_{\mathrm{T}}^{\tau} > 20\,\mathrm{GeV}$, $N_{\mathrm{tracks}}^{\tau} = 1, 3$, $\lvert\eta_\tau\rvert < 2.5$, $\lvert\eta_\tau\rvert \notin [1.37, 1.52]$, all τ^{cand} from same vertex
Analysis level	q_i: 2+2- , $N_{\mathrm{lep}} \leq 1$

Table 6.2: Preselection cuts placed on top of the DAOD cuts shown in table 6.1. They only require certain properties for an object to be a tau candidate and select the desired $3h1l$ and $4h0l$ analysis sub-channels. However, the charge requirement is suspended for the data driven fake estimation procedure described in the sections 6.5 and 6.6. Taken from [8].

The corresponding cutflow histogram is depicted in figure 6.2(a), while figure 6.2(b) and 6.2(c) show the resulting background modeling for some representative observables. Due to their large cross section, the background is dominated by $t\bar{t}$ and single top events. The largest truth-level backgrounds, i.e. Higgs and multiboson processes, only give a low contribution to the background on reconstruction level. Furthermore, a large difference between the background expectation and the recorded data is visible. This leads to the conclusion that there is a significant background contribution exclusively from QCD multijet processes which have not been taken into account yet. Since such events can not

be described by perturbation theory and hence can not be simulated in Monte Carlo simulations, a data driven estimation technique for this background is introduced in the next section.



(a) (b) (c)

Figure 6.2: In (a) the cutflow histogram to table 6.2 is shown. The application of the chosen cuts on the given n-tuples can reduce most of the backgrounds by roughly one order of magnitude. Unfortunately, the loss of some signal events is unavoidable. The remaining plots show the resulting background estimation in comparison with the recorded data on preselection level (b) and after a cut on $\Delta R_{\tau\tau}$ (c). A large discrepancy can be observed.

6.5 Data Driven Estimation of Fake Taus from QCD Multijet Processes

The data driven estimation technique for QCD multijet events used in this analysis has also been developed in [8] and is based on exploiting information from phase space regions with unphysical charge combinations[4]. For the signal processes $hh \to 4\tau$ and $ZZ \to 4\tau$ four correlated taus are expected from which two have positive and the other two negative electric charge. On the other hand, many of the background processes in the 4τ channel produce only two correlated τ-leptons (e.g. $h \to \tau\tau$, $Z \to \tau\tau$ or $t\bar{t}$ decays) and in multijet events, the taus faked by QCD jets are usually not correlated at all. For such completely uncorrelated objects, the tau charges are expected to be distributed with a $1:4:6:4:1$ ratio along the different charge regions, according to the binomial distribution. Table 6.3 shows the observed ratios for the signal processes, expected background and recorded data. The data seems to fulfill the binomial expectation, which justifies the assumption that the events in the unphysical data charge regions are mainly QCD multijet events with four uncorrelated τ-leptons. However, the picture looks different for the background processes. In most of them, two taus directly stemming from the mother particle decay should be correlated while the remaining two are faked and might be uncorrelated with respect to the first two and to each other. The corresponding binomial expectation for the inner charge regions 3+1-, 2+2- and 1+3- would thus read $1:2:1$ and is quite well represented in the observed event yields.
Under the assumption that the QCD multijet events do not differ in their properties along the different charge regions, their number in the 2+2- region can now be estimated from the event yields in the other regions. Therefore, at first the Monte Carlo prediction has to be subtracted from the recorded data in each of the four regions. At this point, it has to be stated that this is only a valid approach if the

[4] For four taus with charge +1 or −1 each they read 4+0-, 3+1-, 1+3- and 0+4-, while 2+2- denotes the physical charge region.

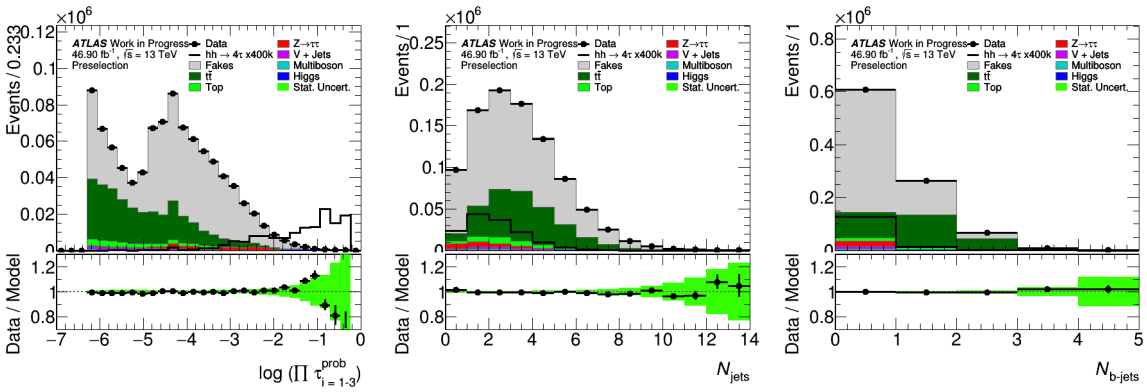
Cut	$hh \to 4\tau$	$ZZ \to 4\tau$	Higgs	$Z \to \tau\tau$	Diboson	V + Jets	$t\bar{t}$	Data
4+0-	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
3+1-	23.5	20.7	82.1	16.6	9.9	8.8	8.5	4.4
2+2-	164.5	84.5	175.4	30.3	17.6	15.9	15.3	6.6
1+3-	24.0	21.2	77.5	15.2	8.9	7.4	8.4	4.1
0+4-	1.0	1.0	0.9	0.8	0.8	0.7	1.0	0.9

Table 6.3: The tau charge ratios for the signal processes, different backgrounds and the recorded data is shown. They are all normalized to their event yields in the 4+0- channel. The $1:4:6:4:1$ ratio expected from the binomial distribution for four uncorrelated objects is only roughly fulfilled for data. The asymmetry towards the positive charge regions is caused by the positive charges of the initially colliding protons in the beams. Taken from [8].

subtracted part is small compared to the overall number of events in the regions. This condition will be of interest in section 6.6 again. Afterwards, the remaining number of events in the outer 4+0- and 0+4- regions is averaged and multiplied by a factor of six (according to the binomial distribution), serving as the first part of the fake estimation for the physical charge region. In a next step, the same events are taken as a starting point again, but now their number is quadruplicated and subtracted from the averaged expectation in the 3+1- and 1+3- regions. The result is then scaled up according to the $1:2:1$ ratio for two correlated objects and added to the first part obtained from the 4+0- and 0+4- regions. This sum is finally used as estimation for the number of QCD multijet events in the 2+2- region.
Figure 6.3 shows the outcome of this procedure. It can be nicely seen that, within the statistical uncertainty, the data is now in good agreement with the background prediction for different observables. The left plot furthermore demonstrates the separation power of the designed tau probability product. While the signal peaks close to a value of zero, the background predominantly accumulates at large negative values. Background processes with two truth τ-leptons like $h \to \tau\tau$ and $Z \to \tau\tau$ peak at intermediate values, as expected. Further plots and more detailed studies concerning the estimation procedure can also be found in [8] and indicate a good data-to-model agreement beyond preselection level as well. Hence, this data driven technique seems to be a good choice for the estimation of the QCD multijet background in the 4τ channel.

6.6 Advanced Data Driven Fake Tau Background Estimation

As described in the previous section, a key requirement for the data driven fake estimation technique to be a valid approximation is that the number of subtracted Monte Carlo simulation events is small compared to the data in the respective unphysical charge regions. Luckily, the distributions of the multijet and Monte Carlo backgrounds are very similar in their shape for most observables, but there is one exception. Looking at figure 6.3(a) again, the tau probability product has been designed in a way to separate multi-truth tau events from fake tau events as best as possible. Thus, events with a high amount of fake taus mainly populate the left half of the plot while there are almost no such events for probability values close to zero. Therefore, it might happen for this phase space region, that the number of Monte Carlo events in a single bin is not small compared to the data, leading to large statistical fluctuations in the fake estimation. Figure 6.4 displays this behavior. The result of the



(a) (b) (c)

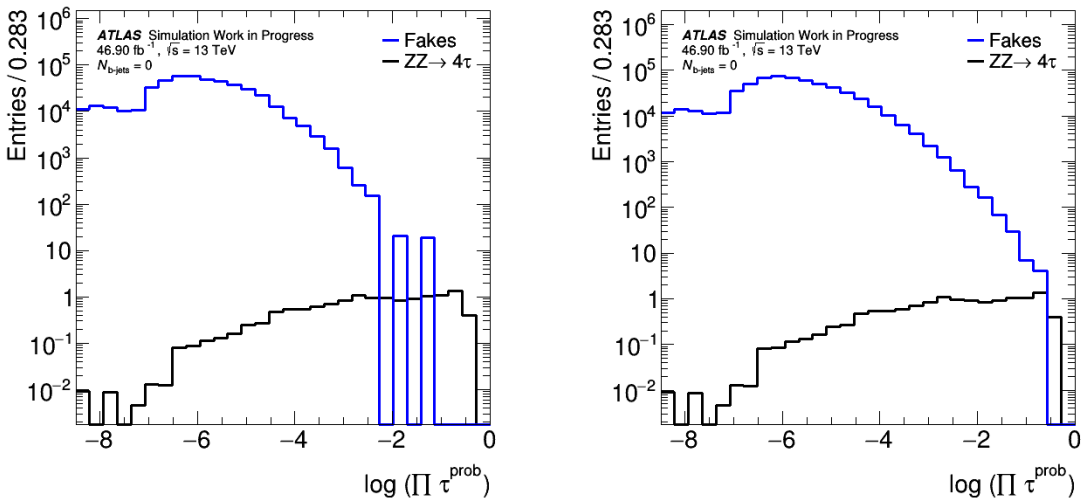
Figure 6.3: The result of the data driven fake estimation from unphysical charge regions is depicted for different observables on preselection level. All three plots show a very good agreement between the totally expected background and the recorded data. In (a), the discrimination power of the tau probability product designed in section 5.2 is clearly visible. Figure (b) can be directly compared to figure 6.2(b) where no QCD multijet fakes have been considered. The right plot (c) shows the distribution of the number of jets containing b-quarks which will be of interest in chapter 7 again.

multijet fake estimation introduced in section 6.5 is compared with a different approach where only Monte Carlo events with three or four, respectively, truth-matched τ-leptons are subtracted during the estimation procedure. This additional requirement on the subtracted events ensures that their number is also small compared to the data events with unphysical charges in the probability product bins close to zero which then leads to a more robust fake estimation. As the fraction of events fulfilling this requirement is very low with respect to the total amount of Monte Carlo events, nearly the whole expected background is then estimated by this data driven technique, as shown in figure 6.5 and 6.6. Nevertheless, it is still modeled very well and a good agreement with the data in the 2+2- region can be observed.

However, at this point it has to be noted that of course the notation is changed. The background denoted as "Fakes" in the plots does not contain pure QCD events anymore, but additionally all events previously taken from Monte Carlo simulations which have at least one tau faked by a jet or light lepton are included. The remaining events summarized in the other background processes are only of the order of single events, their expected yields are shown in table 6.4. Furthermore, the following processes are completely estimated data driven and fall out of the Monte Carlo background:

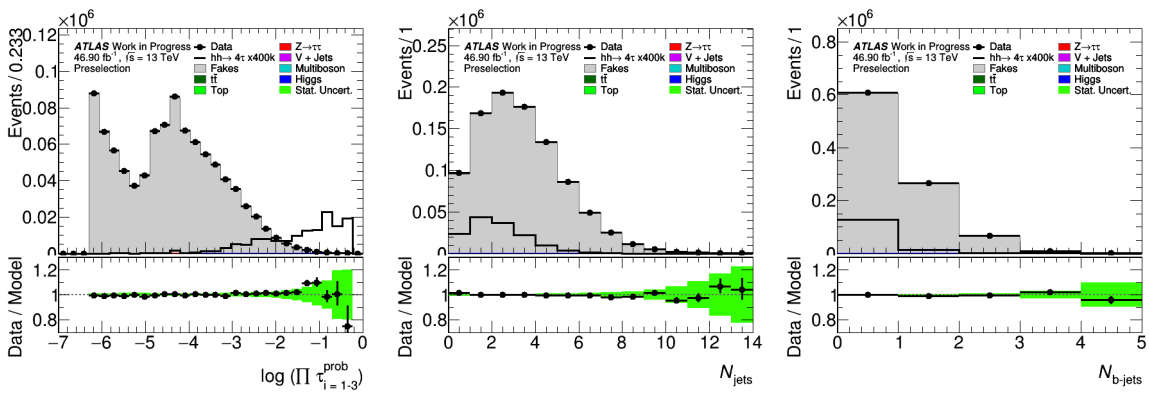
- $Z \rightarrow ee$ + Jets ($3h1l$ & $4h0l$ channel)
- $Z \rightarrow \mu\mu$ + Jets ($3h1l$ & $4h0l$ channel)
- V + Jets ($4h0l$ channel, TChannel processes, cf. tab. A.5)
- $Z \rightarrow \tau\tau$ ($4h0l$ channel)
- Single top ($4h0l$ channel) .

This background estimation will now be used in the further analysis chapters 7 and 8 where a setup allowing for a measurement of the cross section of the $ZZ \rightarrow 4\tau$ process is developed.



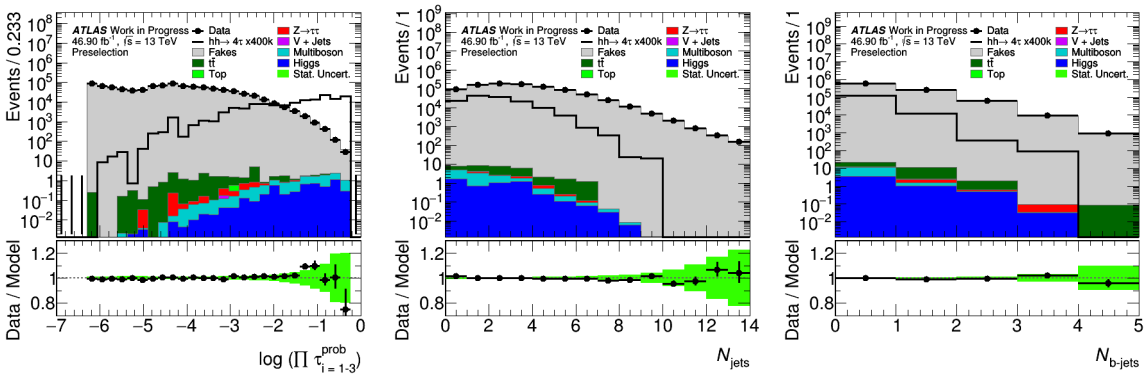
(a) Usual Monte Carlo event subtraction (b) Truth-matched Monte Carlo event subtraction

Figure 6.4: The distribution of the fake tau background and the $ZZ \rightarrow 4\tau$ process in the tau probability product is shown. Additionally, a veto on the existence of b-jets in the event is applied. The left plot (a) shows the result of the estimated multijet background if all Monte Carlo events in each unphysical charge region are subtracted from the recorded data during the estimation procedure. As the number of data events in bins close to zero can be very low, the subtraction might lead to negative fake estimates there. In the right plot (b) an alternative approach is taken, as the requirement of three or four, respectively, truth-matched taus is placed on the Monte Carlo fraction being subtracted. It can be nicely seen that this additional criterion helps to stabilize the fake estimation technique in regions with relatively high tau probability product scores.



(a) (b) (c)

Figure 6.5: The result of the advanced fake tau estimation for the observables from figure 6.3 is shown. Nearly the whole visible background is now predicted by the data driven estimation technique but the agreement between data and model is still very good and differs only slightly from the previous results.



(a) (b) (c)

Figure 6.6: Logarithmic visualization of the results shown in figure 6.5. The remaining contributions with truth-matched taus from the different background processes are visible, although they are very small. All events with not three or four truth-matched taus, respectively, are part of the data driven estimated fake background colored in gray.

Process / Channel	$3h1l$	$4h0l$	Total
V + Jets	0.05 ± 0.05	0	0.05 ± 0.05
Higgs	3.95 ± 0.82	1.12 ± 0.41	5.07 ± 0.92
$Z \to \tau\tau$	1.95 ± 0.50	0	1.95 ± 0.50
Multiboson	0.88 ± 0.09	0.05 ± 0.01	0.93 ± 0.09
$t\bar{t}$	20.31 ± 1.92	0.25 ± 0.25	20.56 ± 1.93
Top	0.30 ± 0.30	0	0.30 ± 0.30
Fakes	381151.19 ± 1797.29	569412.81 ± 2176.78	950564 ± 2823

Table 6.4: Event yields of the different background processes resulting from the advanced data driven fake estimation on preselection level. Only 0.003 % of the background events are still modeled by Monte Carlo simulations, all remaining events are estimated with the data driven technique.

CHAPTER 7

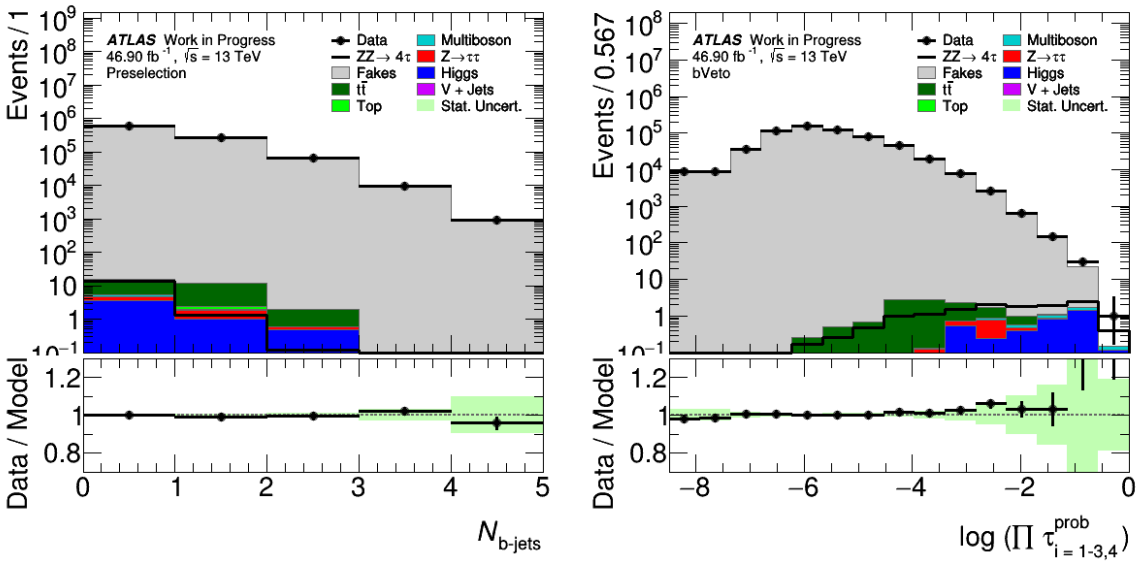
Event Selection in the $ZZ \rightarrow 4\tau$ Channel

In the previous chapters, the tau probability product as a powerful tool for identification of multi-tau events (chap. 5) and a data driven fake tau background estimation method based on using information from unphysical charge regions (chap. 6) have been introduced. For that purpose, a $hh \rightarrow 4\tau$ sample was used to model a 4τ final state. The main aim of this thesis is now to validate these methods on a Standard Model process. Thus, this and the next chapter focus on how to use them to perform a measurement of the $ZZ \rightarrow 4\tau$ cross section. At first, the results of a simple cut-based approach using the tau probability product are shown in section 7.1, while section 7.2 presents a more complex multivariate approach with boosted decision trees in order to further enhance the separation of signal and background events. In section 7.3, the obtained BDT classifiers are finally combined with the information provided by the tau probability product to construct signal and control regions for a profile likelihood fit (cf. chap. 8).

7.1 Cut-Based Approach

As described in section 6.2, the 4τ channel suffers from a large hadronic background whereas the expected number of $ZZ \rightarrow 4\tau$ signal events is very low. Consequently, it must be ensured to not loose too many of these events when trying to design a phase space region from which later the $ZZ \rightarrow 4\tau$ cross section can be obtained. Hence, only two cuts are placed in order to construct such a signal region. Figure 7.1(a) shows the distribution of the number of jets containing b-quarks for signal, background and data. As the ZZ decay does not produce any b-quarks, one would expect the signal process to mainly populate the first bin of the plot. This behavior can indeed be observed. In contrast to that, especially $t\bar{t}$ events with three or four truth-matched tau candidates and the fake-tau background produce final states with a number of b-jets unequal to zero. Therefore, a requirement of $N_{\text{b-jets}} = 0$, referred to as b-veto in the following, can already reject a large fraction of the expected background events while hardly any signal events are lost. In figure 7.1(b) the distribution of the tau probability product with an b-veto applied is depicted. Similar to the $hh \rightarrow 4\tau$ process studied in chapter 6, it also yields a high separation power between the fake background and the $ZZ \rightarrow 4\tau$ signal, although the overlap of signal and the background with truth-matched taus is larger compared to the di-Higgs process.

The resulting events yields for different $\log\left(\prod \tau^{\text{prob}}_{\text{i=1-3,4}}\right)$ cut values are summarized in table 7.1 while



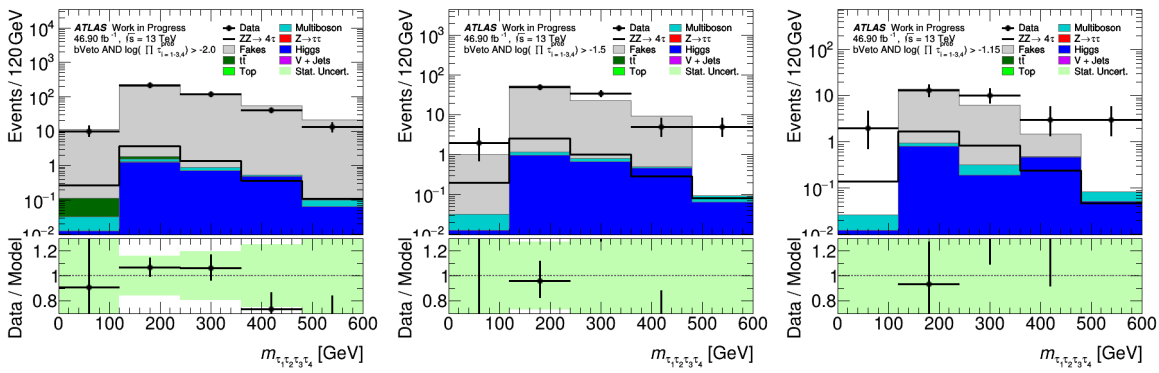
(a) (b)

Figure 7.1: Distributions of the different cut variables used to enhance the signal-to-background separation in the $ZZ \rightarrow 4\tau$ channel. Subfigure (a) shows the distribution of the number of b-jets. While the majority of signal events does not produce any b-jets, the fake-tau background has a quite large fraction of events in the region $N_{\text{b-jets}} \neq 0$. In (b), the tau probability product is depicted after placing the requirement of $N_{\text{b-jets}} = 0$. A good separation between the fake-tau enriched background and the $ZZ \rightarrow 4\tau$ signal can be observed.

figure 7.2 exemplary shows the distribution of the invariant mass of the 4τ system after the application of the cuts. The effect of the b-veto can be determined by comparison with table 6.4 where the yields on preselection level were given. The dominating fake background can be reduced by roughly 36 % and also the largest Monte Carlo simulated process, $t\bar{t}$ can be halved, although it is already very small in terms of absolute events compared to the fakes. The further columns show the impact of different cuts on the tau probability product. As it has been designed for that purpose, the rejection against the fake-tau background is very high and with increasing cut values, the remaining background mainly consists of only Higgs and multiboson events in addition. This observation is visible in figure 7.2 as well and matches the results of the truth-level study performed in [8] which was shortly mentioned in section 6.3. The depicted 4τ invariant mass distributions show a peak at a value slightly below 200 GeV. As the Z boson has a mass of around 91 GeV, one would expect the di-Z invariant mass (which is equal to the mass of the 4τ system) to peak close to 180 GeV, if the Z bosons are produced at rest. But since the neutrinos from the tau decays can not be reconstructed in the detector, the visible invariant mass is expected to be shifted to a lower value. This behavior can not be observed directly in the plots since the binning had to be chosen quite coarse in order to have enough statistics in each bin, but the result roughly confirms the expectation. Albeit the achievable significance $S/\sqrt{S+B}$ rises with the choice of higher cut values on $\log\left(\prod \tau^{\text{prob}}_{\text{i=1-3,4}}\right)$ (cf. tab. 7.1 again), the loss of signal events is unavoidable and limits the performance that can be achieved with this cut-based event selection. Therefore, a more complex approach with boosted decision trees is presented in the next section which results in a better signal-to-background separation.

Process / Cut	b-veto	$\log\left(\prod \tau^{\text{prob}}_{\text{i=1-3,4}}\right) > -2.0$	$\log\left(\prod \tau^{\text{prob}}_{\text{i=1-3,4}}\right) > -1.5$	$\log\left(\prod \tau^{\text{prob}}_{\text{i=1-3,4}}\right) > -1.15$
$ZZ \to 4\tau$	13.36 ± 0.32	5.85 ± 0.27	4.20 ± 0.25	2.97 ± 0.22
V + Jets	0.05 ± 0.05	0	0	0
Higgs	3.54 ± 0.82	2.55 ± 0.71	2.22 ± 0.71	1.54 ± 0.58
$Z \to \tau\tau$	0.97 ± 0.40	0.08 ± 0.06	0	0
Multiboson	0.74 ± 0.08	0.54 ± 0.06	0.46 ± 0.06	0.37 ± 0.05
$t\bar{t}$	9.68 ± 1.30	0.33 ± 0.17	0	0
Top	0	0	0	0
Fakes	606289 ± 2236	405 ± 44	82.83 ± 20.52	19.95 ± 9.90
Data	607814 ± 780	405 ± 20	99 ± 9.95	31 ± 5.57
$S/\sqrt{S+B}$	0.017	0.287	0.443	0.635

Table 7.1: Event yields of the $ZZ \to 4\tau$ signal, the background processes and recorded data after the cumulative application of a b-veto and different $\log\left(\prod \tau^{\text{prob}}_{\text{i=1-3,4}}\right)$ cuts. Cutting on $\log\left(\prod \tau^{\text{prob}}_{\text{i=1-3,4}}\right)$ reduces the fake-enriched background drastically and the larger the cut value is, the higher is the background rejection which can be achieved. Even after the application of the cuts there is good agreement between background prediction and data within the statistical uncertainty. The last row additionally shows the resulting significance $S/\sqrt{S+B}$ after each cut.



(a) b-Veto, $\log\left(\prod \tau^{\text{prob}}_{\text{i=1-3,4}}\right) > -2.0$ (b) b-Veto, $\log\left(\prod \tau^{\text{prob}}_{\text{i=1-3,4}}\right) > -1.5$ (c) b-Veto, $\log\left(\prod \tau^{\text{prob}}_{\text{i=1-3,4}}\right) > -1.15$

Figure 7.2: Invariant mass distribution of the 4τ system after the application of a b-veto and different cuts on $\log\left(\prod \tau^{\text{prob}}_{\text{i=1-3,4}}\right)$. As expected, the number of background events decreases with an increasing cut value and only the fakes estimated from data as well as Higgs and multiboson events yield a non-vanishing contribution to the remaining background.

7.2 Multivariate Approach using Boosted Decision Trees

Beside neural networks, boosted decision trees are a common tool in high energy physics analyses to classify events either as signal- or background-like. Already a forest of a few simple trees can lead to a considerable improvement in the separation between these two event classes. In this analysis, two independent forests are trained and evaluated to discriminate the $ZZ \rightarrow 4\tau$ signal from the expected background. The training setup and the resulting separation performance are described in the subsequent subsections.

7.2.1 Training Samples and Variables

Although nearly the whole background for the 4τ channel is estimated with a data driven method, it still consists of different underlying physical processes. Taking a look back at figure 6.3, the contribution of heavy particle decays to the total background is roughly 33 % and is dominated by $t\bar{t}$ or single top quark events. The remaining fraction originates from QCD multijet processes which then are misidentified as τ-leptons. Since the jets in the latter events might be completely uncorrelated and are exclusively produced by the strong interaction, their kinematic and other physical properties are expected to differ from those of the tau candidates being produced in heavy particle decays. Thus, two independent forests of decision trees are set up and trained in order to separate $ZZ \rightarrow 4\tau$ events from the multijet and top quark-enriched background, respectively. Each forest is build up of 400 decision trees with a maximum depth of 3 and minimum node size of 5 %. For each single tree, *AdaBoost* and *GradientBoost*, which were introduced in section 4.3.2, are used for boosting. The available events are split based on their event number into training and test events. The training sample consists of those events with an even event number, while events with an odd number are taken for testing. For the training procedure, they all are reweighted to the absolute value of their nominal weight. This ensures that events with negative weights do not affect the training in a detrimental way. Table 7.2 summarizes these parameters for the BDT training. Furthermore, the samples for the training have to be chosen. The decision tree for separation of signal from QCD multijet events (referred to as fake or multijet BDT with classifiers *BDT (Fakes)* using *AdaBoost* and *BDTG (Fakes)* using *GradientBoost*) is trained on data events from the unphysical charge regions (i.e. 4+0-, 3+1-, 1+3- and 0+4-) as they directly represent the properties of the multijet fakes. The second decision tree (referred to Monte Carlo BDT with classifiers *BDT (MC)* and *BDTG (MC)*) uses all background events in the 2+2- region generated by Monte Carlo simulations as training input, i.e. those samples specified in appendix A, except for the $hh \rightarrow 4\tau$ sample and the fraction of events in the $\ell\ell\ell\ell$ sample serving as $ZZ \rightarrow 4\tau$ signal (cf. section 6.2). Additionally, a b-veto is placed on all training events. The idea of this requirement has already been motivated in section 7.1 and ensures that the decision trees focus on the determination of more complex event properties and correlations between single observables. The resulting numbers of training and test events for both trees are summarized in table 7.3.
After all these parameters are fixed, the training variables for the decision trees need to be selected. A careful optimization has been performed and the results are given below. At first, the whole BDT setup is intended to be independent from the tau identification method introduced in section 5.2. This ensures that the classifier outputs provided by the decision trees are orthogonal to the $\log\left(\prod \tau^{\text{prob}}_{\text{i=1-3,4}}\right)$ observable and can be later combined with it for the measurement of the $ZZ \rightarrow 4\tau$ cross section. Therefore, only kinematic variables like the p_{T} of tau candidates, invariant masses of ditau systems or distances in ϕ, η or R between reconstructed objects are considered for the training. Any tau

N_{trees}	400
Tree depth	3
Min. node size	5 %
Boost types	*AdaBoost*, *GradientBoost*
Training events	N_{event} even
Test events	N_{event} odd
Event weight	|nominal weight|
Transformations	I, P, G, D

Table 7.2: General training parameters of the boosted decision trees used for discrimination of the $ZZ \rightarrow 4\tau$ signal against its background. They have been optimized based on the performance of the *GradientBoost* which will turn out to yield a better separation than *AdaBoost* does (cf. sec. 7.2.3 and 7.2.4). The transformation abbreviations mean ***I**dentity*, ***P**CA (Principal Component Analysis)*, ***G**aussianisation* and ***D**ecorrelation*.

track-related information has to be left out since it is already used in the tau identification procedure, albeit many of these variables would yield a higher separation power than the majority of the kinematic ones. The training variable optimization was started from the same set of more than 170 variables for both trees, from which at the end the subsets yielding the best separation performance on each training sample have been chosen. These subsets include 45 variables for the Monte Carlo BDT and 34 for the fake BDT. Before the training process, these variables are transformed by the TMVA `Factory` (cf. sec. 4.3.1) according to the transformations set in the training parameters. In this analysis, the transformations are an identity transformation followed by the principal component analysis (PCA), a Gaussian transformation and finally a decorrelation transformation. Some exemplary input variable distributions for both BDTs resulting from the application of all transformations are depicted in the figures 7.3 and 7.4. The largest separation power in the shown variables for the multijet BDT is provided by the transformed MET significance[1] (fig. 7.3, top left), but also $\Delta\eta_{\tau_1\tau_3}$ indicates a different distribution for signal and background events. The other distributions look quite similar for both classes of events. However, this does not imply that the BDT performs bad because in addition to the 1-dimensional distributions plotted here, it also considers two- and higher dimensional correlations between the different variables. For the Monte Carlo BDT, this feature seems to be even more important, as figure 7.4 reveals a large overlap between the signal and background histograms for almost all variables. This behavior might be somehow expected since these processes should be more similar in their physical properties than the signal compared to QCD multijet events produced by the strong interaction. Hence, the input variable plots give already a first hint towards possible performance differences between both decision trees which are covered in the next sections. For the sake of completeness, the remaining training variable distributions can be found in appendix B.

7.2.2 Training Results

TMVA itself already provides a lot of different output information which allows for the optimization of the BDT training. Using them is important, because any improvement of the training procedure might directly result in a better signal-to-background separation in the classifier output. The information mainly considered for that purpose in this analysis is the ranking of input variables created by TMVA in

[1] The MET significance is a measure for the probability that an observed MET value is consistent with a fluctuation from zero because of detector-related limitations like finite measurement resolution [68].

Monte Carlo BDT	Training	$N_{\text{sig}} = 6893$, $N_{\text{backg}} = 481316$
	Testing	$N_{\text{sig}} = 6645$, $N_{\text{backg}} = 481574$
Fake BDT	Training	$N_{\text{sig}} = 6893$, $N_{\text{backg}} = 562830$
	Testing	$N_{\text{sig}} = 6645$, $N_{\text{backg}} = 562383$

Table 7.3: Number of training and test events of the boosted decision trees used for discrimination of the $ZZ \rightarrow 4\tau$ signal against its background.

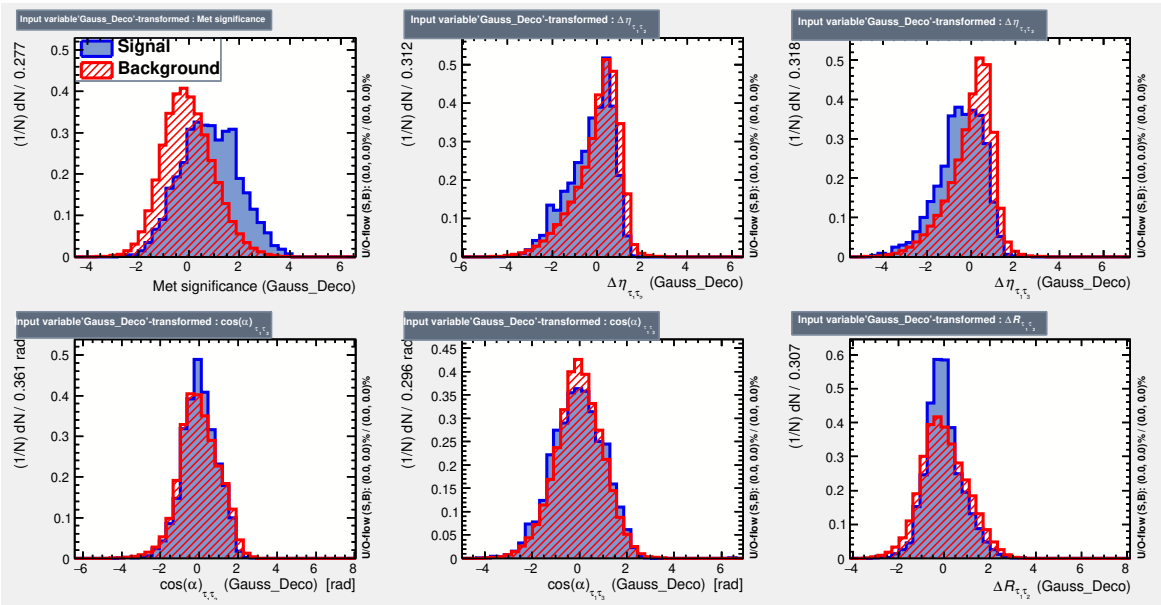


Figure 7.3: Exemplary input variable distributions after application of all transformations mentioned in table 7.2 of the decision tree trained for discrimination of the $ZZ \rightarrow 4\tau$ signal against QCD multijet events. The signal histograms are colored in blue while the background is depicted in red. Especially the plots in the top left and top right corner show a visible separation between both classes of events.

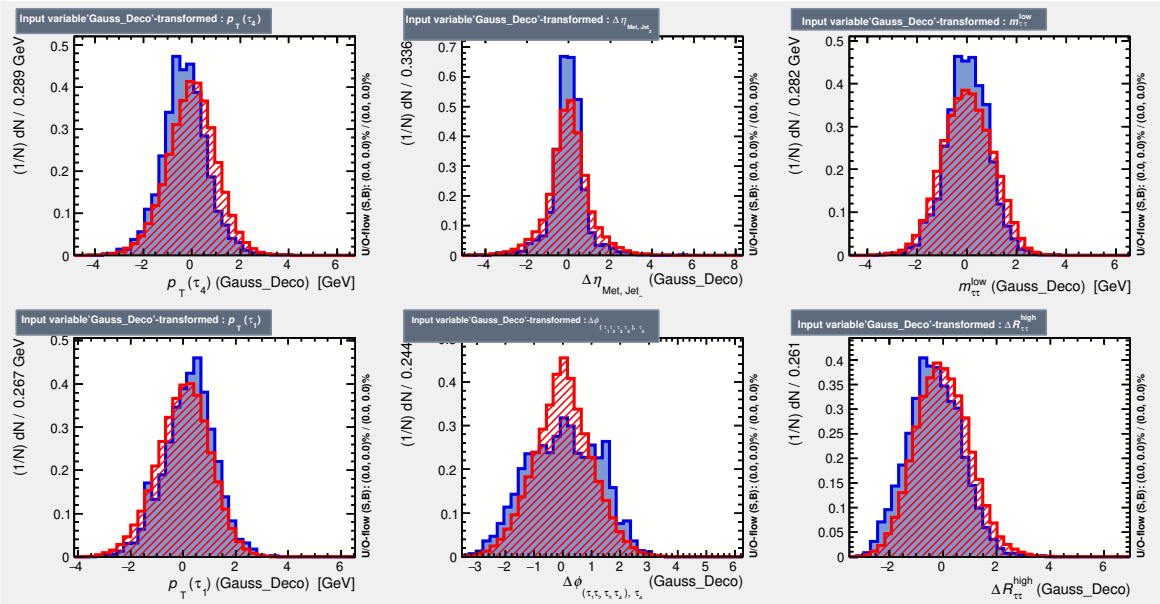


Figure 7.4: Exemplary input variable distributions after application of all transformations mentioned in table 7.2 of the decision tree trained for discrimination of the $ZZ \rightarrow 4\tau$ signal against Monte Carlo background events. The color code is equal to figure 7.3, but a comparison with this figure shows a generally larger overlap between signal and background distributions for the Monte Carlo BDT.

combination with the area under the ROC (*Receiver Operating Characteristic*) curve which visualizes the resulting background rejection as a function of signal efficiency. For each trained classifier, TMVA evaluates the importance of each single input variable with respect to all splitting nodes and trees and gives a score to it. These scores can not be interpreted numerically in a direct way, but the higher the score, the more often the variable is used for the separation of signal and background events in the training process. Thus, starting from a very large set of variables, the training efficiency can be improved by basically removing one variable after another from the training until the performance does not increase anymore because the tree gets to less information. This performance is evaluated by the area under the ROC curve (AUC) which is also directly calculated after the training step. At this point it is important to further check for the effect of overtraining to ensure that the performance achieved on the training sample is similar to the one on the test sample and represents the whole dataset.

The classifier-specific top-ranked variables for both decision trees are summarized in table 7.4 and 7.5, while the full rankings including all variables of the final selection are listed in appendix C. For the Monte Carlo BDT, it is found that the sum of transverse momenta of the tau candidates considering only the $3h1l$ channel[2] and the missing transverse momentum play the most important role in the training process when using *AdaBoost*, followed by the difference in η between the MET and the jet with the third-highest p_T[3] (if existing). In the case of using *GradientBoost*, the variables with the

[2] The fourth tau candidate τ_4 refers to the tau with the lowest τ^{jetBDT} score in the $4h0l$ channel, whereas it is defined as the charged lepton for events in the $3h1l$ channel.

[3] In contrast to the tau candidates, which are sorted with respect to their τ^{jetBDT} score, the reconstructed jets are numbered according to their transverse momentum, i.e. jet_1 is the leading-p_T jet, jet_2 the subleading-p_T one, etc.

highest importance are distances in ϕ between the leading jet and the third or fourth tau candidate, respectively, as well as $\Delta\phi$ of both ditau systems with falsely assigned tau candidates. The training of the QCD multijet BDT mainly benefits from the different distributions in the transverse missing momentum and the MET significance, which could already be expected from e.g. the top left plot in figure 7.3. Similar to the Monte Carlo BDT, the distance in ϕ between the leading jet and the fourth tau candidates is an important separation variable here, too.

Rank	Variable	Importance
1	$\sum p_{\mathrm{T}}(\mathrm{lep}_{1-4}, 3h1l)$	3.410×10^{-2}
2	$p_{\mathrm{T}}^{\mathrm{Met}}$	3.169×10^{-2}
3	$\Delta\eta_{\mathrm{Met,\ Jet_2}}$	3.025×10^{-2}
4	$\Delta\phi_{\mathrm{min}}((\tau_2\tau_4), \mathrm{Met})$	2.863×10^{-2}
5	$\Delta R_{\tau\tau}^{\mathrm{high}}$	2.735×10^{-2}
6	$\Delta\phi_{\tau_1\tau_3}$	2.595×10^{-2}
7	$\Delta\phi_{\mathrm{Met},\tau_4}$	2.567×10^{-2}
8	$m_{\tau\tau}^{\mathrm{low}}$	2.548×10^{-2}
9	$\Delta\phi_{\mathrm{min}}((\tau_1\tau_3), \mathrm{Met})$	2.542×10^{-2}
10	Met significance	2.510×10^{-2}

(a) BDT (MC)

Rank	Variable	Importance
1	$\Delta\phi_{\mathrm{Jet}_1,\tau_4}$	2.908×10^{-2}
2	$\Delta\phi_{(\tau_1\tau_3),(\tau_2\tau_4)}$	2.856×10^{-2}
3	$\Delta\phi_{\mathrm{Jet}_1,\tau_3}$	2.754×10^{-2}
4	$\Delta R_{\tau\tau}^{\mathrm{low}}$	2.678×10^{-2}
5	$\Delta\phi_{(\tau_1\tau_2\tau_3\tau_4),\tau_4}$	2.663×10^{-2}
6	$\Delta\eta_{\mathrm{Met},\tau_4}$	2.640×10^{-2}
7	$\Delta\phi_{\tau_1\tau_3}$	2.626×10^{-2}
8	$\Delta\eta_{\tau_3\tau_4}$	2.580×10^{-2}
9	$\Delta\phi_{\mathrm{Met},\tau_4}$	2.578×10^{-2}
10	$\Delta\eta_{\mathrm{Jet}_1,\tau_3}$	2.558×10^{-2}

(b) BDTG (MC)

Table 7.4: Classifier-specific top-10 ranked input variables for the training of the Monte Carlo BDT.

Rank	Variable	Importance
1	$p_{\mathrm{T}}^{\mathrm{Met}}$	5.837×10^{-2}
2	Met significance	5.585×10^{-2}
3	$\Delta\phi_{\mathrm{Jet}_1,\tau_4}$	4.507×10^{-2}
4	$\sum p_{\mathrm{T}}(\mathrm{lep}_{1-4}, 3h1l)$	4.455×10^{-2}
5	$\Delta\eta_{\tau_1\tau_3}$	3.535×10^{-2}
6	$\Delta R_{\tau\tau}^{\mathrm{high}}$	3.396×10^{-2}
7	$\Delta\phi_{\mathrm{Met},(\tau_1\tau_2\tau_3\tau_4)}$	3.293×10^{-2}
8	$m_{\tau\tau}^{\mathrm{low}}$	3.284×10^{-2}
9	$p_{\mathrm{T}}(\tau_1\tau_2\tau_3\tau_4)$	3.273×10^{-2}
10	$\Delta\phi_{\mathrm{Met},\tau_4}$	3.246×10^{-2}

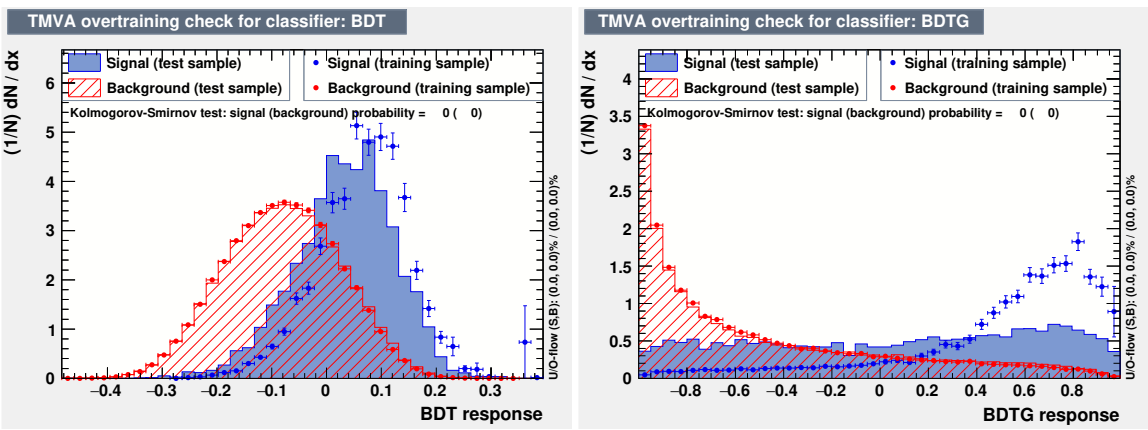
(a) BDT (Fakes)

Rank	Variable	Importance
1	$p_{\mathrm{T}}^{\mathrm{Met}}$	4.600×10^{-2}
2	Met significance	4.569×10^{-2}
3	$\Delta\eta_{\tau_1\tau_3}$	3.506×10^{-2}
4	$\Delta\phi_{\mathrm{Jet}_1,\tau_4}$	3.457×10^{-2}
5	$\Delta\eta_{\mathrm{Met},\tau_4}$	3.450×10^{-2}
6	$\Delta R_{\tau\tau,\mathrm{false}}^{\mathrm{low}}$	3.443×10^{-2}
7	$\Delta\eta_{\tau_3\tau_4}$	3.389×10^{-2}
8	$\cos(\alpha)_{\tau_1\tau_3}$	3.378×10^{-2}
9	$\Delta\phi_{\mathrm{Met},\tau_4}$	3.368×10^{-2}
10	$\Delta R_{\tau\tau}^{\mathrm{high}}$	3.285×10^{-2}

(b) BDTG (Fakes)

Table 7.5: Classifier-specific top-10 ranked input variables for the training of the QCD multijet BDT.

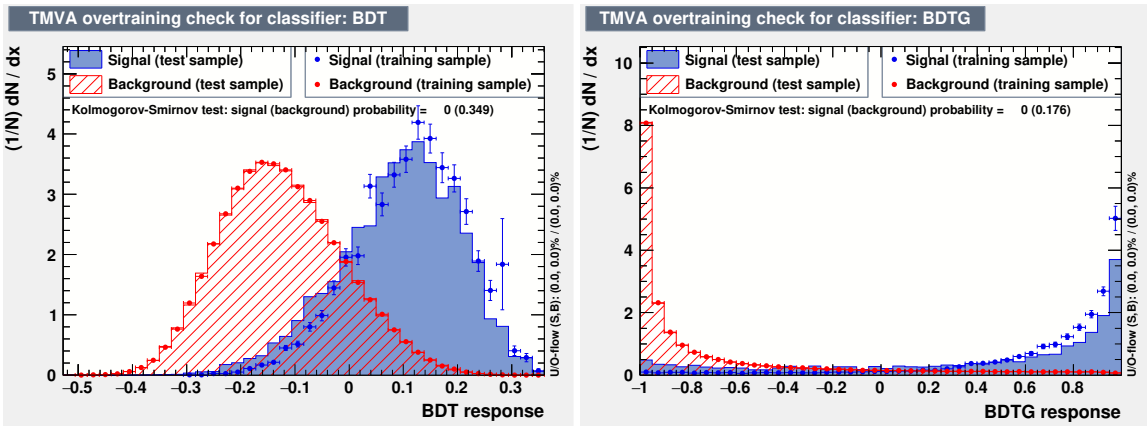
The results of the overtraining check are visualized in figure 7.5 and figure 7.6, table 7.6 additionally displays the corresponding numerical values. As explained above, in the ideal case the histograms for both the training and the test sample should look the same for signal and background, respectively. While this is indeed visible for the background in all plots, there are some deviations for the signal events and the mismatching is a lot larger for the Monte Carlo BDT than for the multijet BDT. The main reason for the occurrence of this effect is assumed to be the low number of signal events for the training procedure, which was limited to approximately 13 500 whereas a few 100 000 events were available for the background. It has been mentioned in chapter 4, that in such a case the decision tree might not reveal the connections between the input variables but "memorizes" single events. Of course this is an extreme case, however it might convey the idea of what is going on there. This problem can usually be tackled by making the decision tree more simple in its structure. Since the tree depth is already quite low (depth 3), the complexity has mainly been managed via the minimum node size parameter, which was varied between 2 % and 20 %. Finally, a value of 5 % was chosen (cf. tab. 7.2). In general, it has been found that the effect of overtraining decreases with a rising minimum node size. On the other hand, the increase of this parameter has negative consequences for the separation between signal and background events in the classifier output. This behavior can be seen very well in appendix D, where the outcome of the overtraining check for all classifiers and different sets of hyperparameters is visualized. Consequently, a compromise had to be found between these effects and the classification performance whose results are discussed in the next subsections.



(a) BDT (MC)

(b) BDTG (MC)

Figure 7.5: BDT classifier output distributions for the different combinations of signal/background and training/test sample for the Monte Carlo BDT (overtraining check). Ideally, both histograms of one color should be identical. While this is well fulfilled for the background, the signal shows differences between training and test sample for both classifiers.



(a) BDT (Fakes)

(b) BDTG (Fakes)

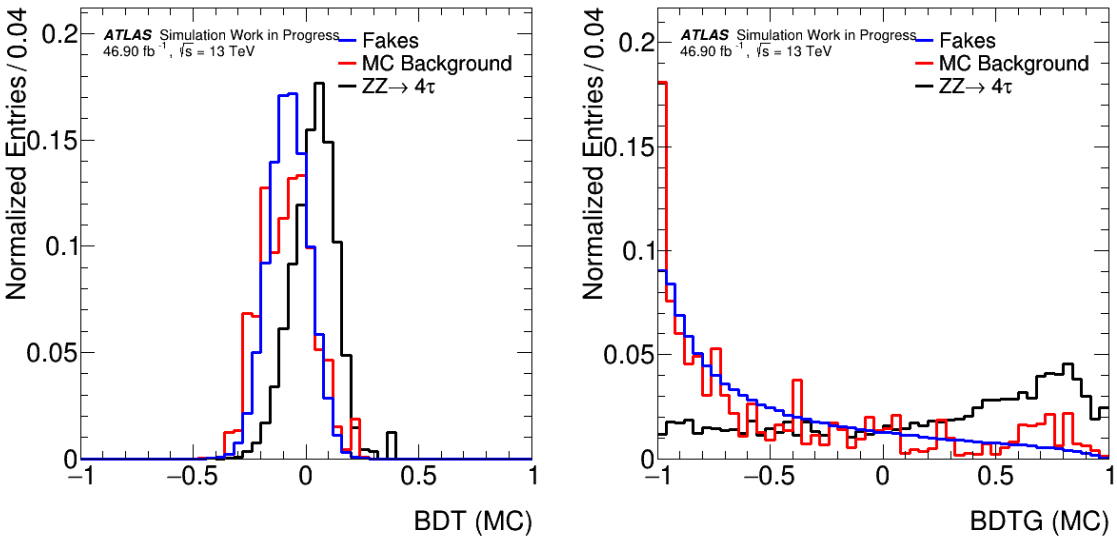
Figure 7.6: BDT classifier output distributions for the different combinations of signal/background and training/test sample for the QCD multijet BDT (overtraining check). Compared to the Monte Carlo BDT (fig. 7.5), the agreement in the background is very good again, and also the signal histograms for the test sample match the ones of the training sample much better. Additionally, the general overlap between signal and background is smaller here, i.e. the multijet BDT performs better than the Monte Carlo BDT does.

Classifier	$\varepsilon_{\text{sig}}^{\text{test}}$ ($\varepsilon_{\text{sig}}^{\text{train}}$)		
	$@\varepsilon_{\text{backg}} = 0.01$	$@\varepsilon_{\text{backg}} = 0.10$	$@\varepsilon_{\text{backg}} = 0.30$
BDT (MC)	0.085 (0.096)	0.417 (0.438)	0.747 (0.754)
BDTG (MC)	0.065 (0.080)	0.364 (0.403)	0.678 (0.701)
BDT (Fakes)	0.323 (0.321)	0.735 (0.734)	0.907 (0.907)
BDTG (Fakes)	0.310 (0.313)	0.720 (0.720)	0.902 (0.902)

Table 7.6: Signal efficiency comparison between training and test sample at fixed background efficiencies for the different BDT classifiers (overtraining check). The numbers for the multijet fake BDT show better agreement and thus less overtraining effects than those for the Monte Carlo BDT.

7.2.3 Monte Carlo BDT Performance

The Monte Carlo BDT has been designed to separate the $ZZ \rightarrow 4\tau$ signal from the background events with truth-matched τ-leptons as well as from the top-enriched Monte Carlo fake tau background. Its classifier output distributions are depicted in figure 7.7. They show a reasonable separation between signal and both type of background events. Monte Carlo background events with at least three truth-matched taus are better separated from the signal as the fake tau events. Nevertheless, since the top-enriched events faking τ-leptons are estimated data driven and therefore also part of the background labeled as "Fakes" here, the BDT provides also some rejection power against the blue histogram. The large fluctuations in the red Monte Carlo background histograms can be explained by the low statistics in those samples. Remembering the results of table 6.4, only approximately 30 background events are not estimated using the data driven method but are directly taken from Monte Carlo simulations. The use of the full run 2 dataset with an integrated luminosity of 139 fb^{-1}

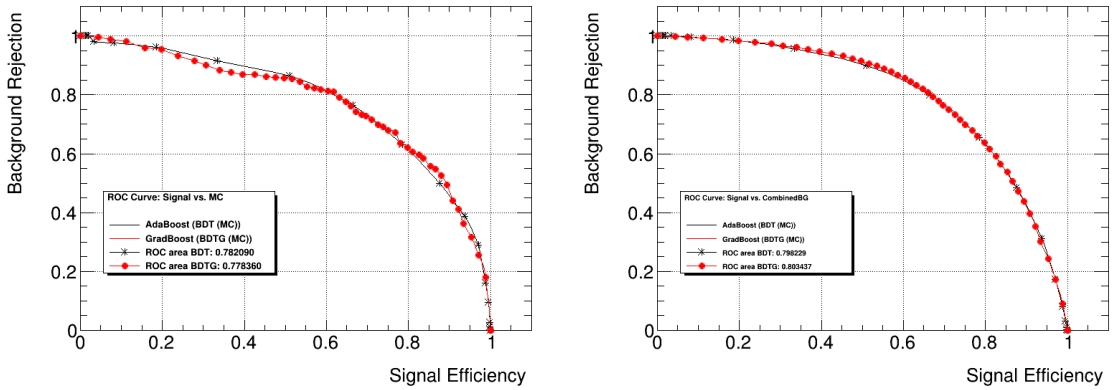


(a) BDT (MC)

(b) BDTG (MC)

Figure 7.7: Normalized classifier output distributions of the Monte Carlo BDT, split up into their performance on the data driven estimated fakes and the remaining background taken from Monte Carlo simulations. In both plots the separation of the Monte Carlo events colored in red is slightly better than those of the fakes depicted in blue. The large statistical fluctuations in the red distributions are caused by the low number of events in the Monte Carlo samples which is in the order of 30 (cf. tab. 6.4).

would probably further enhance the BDT performance at this point. In a next step, a ROC curve can be directly constructed from the obtained classifier distributions. For that purpose, each of them is scanned from its minimum to maximum value with a fixed cut step width of 0.04 (corresponding to 50 data points for a classifier range from -1 to 1). At each of these points, the requirement of a classifier score higher than the current cut value is placed and the resulting number of signal and background events is calculated. The ROC curve then displays the resulting background rejection as a function of the remaining signal efficiency. For the Monte Carlo BDT it is visualized in figure 7.8. In the left plot, the performance is shown for the classifiers being evaluated only on the background events taken from Monte Carlo simulations (red histograms in fig. 7.7), as this is what they have been trained for. In that case, AUC values of 0.782 for the *AdaBoost* and 0.778 for the *GradientBoost* are obtained. These values are in agreement with the results in table 7.6 where also the *AdaBoost* classifier yielded higher signal efficiencies at fixed background rejections than the one which uses *GradientBoost*. However, when being evaluated on the total background (right plot in fig. 7.8), the *BDTG (MC)* classifier performs better than the *BDT (MC)* does. The comparison of the AUC values reads 0.798 (*BDT (MC)*) vs. 0.803 (*BDTG (MC)*). This leads to the assumption that the *GradientBoost* has a much higher rejection power against especially the multijet fakes which dominate the total background, compared to the *AdaBoost*. Hence, only the *BDTG (MC)* classifier is used in the further analysis. Nevertheless, the AUC values of the other tested hyperparameter setups are summarized in table 7.7 for comparison. As already expected from the observations made in the previous section, the performance drops with decreasing tree depth and increasing minimum node size.



(a) Evaluation on the Monte Carlo background only.

(b) Evaluation on the total background.

Figure 7.8: ROC curves evaluating the performance of the Monte Carlo BDT classifiers on different backgrounds. The left plot (a) shows a larger fluctuation, especially in the *GradientBoost* classifier, which is again caused by the low absolute number of events in this background sample. The curves in the right plot (b) are more stable since there the evaluation is performed on the total background. The area under the curve, going towards 1 for an ideal classifier, is around 0.8 for both classifiers in both plots.

Min. Node Size	Tree depth		
	2	3	5
2 %	0.775	0.794	0.816
5 %	0.774	**0.798**	0.807
10 %	0.768	0.785	0.785
20 %	0.735	0.740	0.740

(a) BDT (MC)

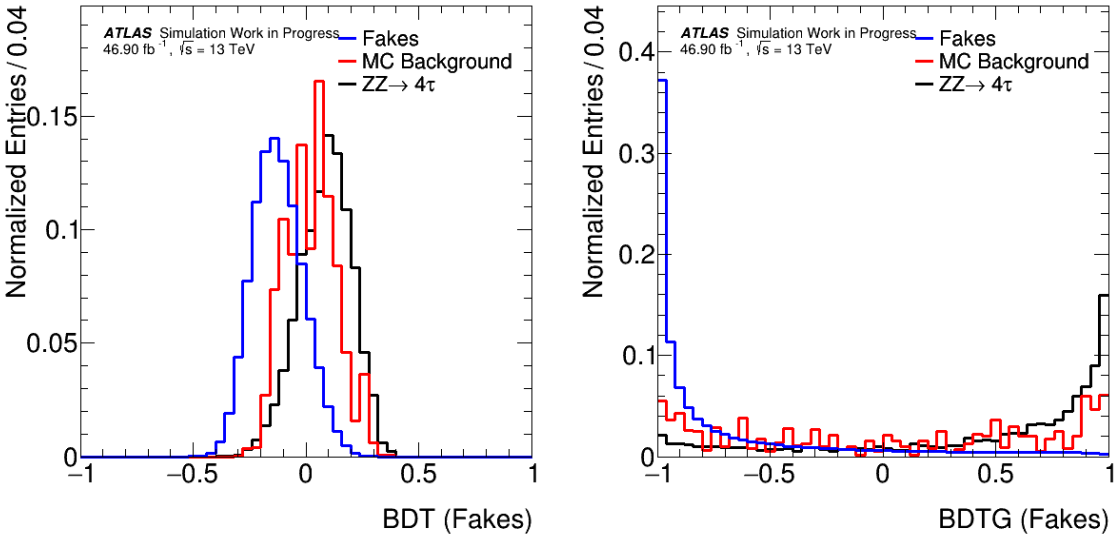
Min. Node Size	Tree depth		
	2	3	5
2 %	0.792	0.807	0.809
5 %	0.791	**0.803**	0.803
10 %	0.774	0.790	0.791
20 %	0.741	0.747	0.747

(b) BDTG (MC)

Table 7.7: AUC values of the *BDT (MC)* and *BDTG (MC)* classifiers evaluated on the total background for different hyperparameter setups. The values get worse with decreasing tree depth and increasing minimum node size. The result for the finally chosen setup is highlighted in bold font.

7.2.4 QCD Multijet BDT Performance

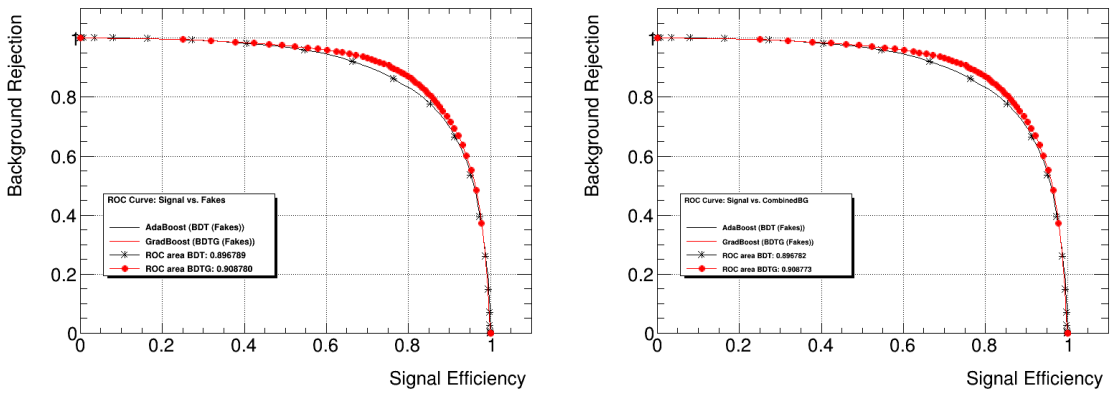
In contrast to the Monte Carlo BDT, the main task of the QCD multijet BDT is the separation of the expected signal from the fake tau events produced exclusively by the strong interaction. As larger differences in the physical properties between such events are assumed compared to the Monte Carlo background, the multijet BDT should also yield an even better separation performance than the Monte Carlo BDT does. The obtained results are depicted in figure 7.9. Indeed, both event classes are separated very well and the BDT power becomes particularly visible in the *BDTG (Fakes)* classifier. Albeit the distributions are quite flat with respect to the Monte Carlo background colored in red, i.e. the overlap of this sample with the signal histogram is large, this behavior is no disadvantage because those events can be handled with the BDT discussed in the previous section. Consequently, the ROC curve, which can be found in figure 7.10, looks also much better compared to the Monte Carlo BDT.



(a) BDT (Fakes)

(b) BDTG (Fakes)

Figure 7.9: Normalized classifier output distributions of the QCD multijet BDT, split up into their performance on the data driven estimated fakes and the remaining background taken from Monte Carlo simulations. While the black and red histograms have a large overlap, the fakes in blue are very well separated from the signal events.



(a) Evaluation on the data driven background only.

(b) Evaluation on the total background.

Figure 7.10: ROC curves evaluating the performance of the QCD multijet BDT classifiers on different backgrounds. As the contribution of the fakes to the total background is extremely close to 100 %, hardly any difference between both plots can be seen. The area under the curve is of around 0.9 for the two classifiers and therefore much better compared to the Monte Carlo BDT, as expected.

Here, it does not make any large difference for the result whether the *BDT (Fakes)* and *BDTG (Fakes)* classifiers are only evaluated on the fakes or on the total background, as both are approximately equal in absolute numbers. The AUC values for the evaluation against all background events are given by 0.897 (*BDT (Fakes)*) and 0.909 (*BDTG (Fakes)*). Again, the procedure has been repeated for different hyperparameter setups and the results are summarized in table 7.8. A similar behavior as for the Monte Carlo BDT is observed. For a better overview, the values obtained from the different decision trees for the finally chosen setup are again summarized in table 7.9. In contrast to figure 7.8, where both classifiers yielded a similar performance, the *GradientBoost* now clearly outperforms the *AdaptiveBoost* in the separation of signal from the total background. Therefore, again the *BDTG (Fakes)* classifier is chosen for the further analysis procedure and is used together with the *BDTG (MC)* and $\log\left(\prod \tau^{\text{prob}}_{\text{i=1-3,4}}\right)$ variables in the next section 7.3 in order to construct different phase space regions for the profile likelihood fit carried out in chapter 8.

Min. Node Size	Tree depth		
	2	3	5
2 %	0.882	0.900	0.918
5 %	0.883	**0.897**	0.911
10 %	0.879	0.892	0.898
20 %	0.851	0.852	0.852

(a) BDT (Fakes)

Min. Node Size	Tree depth		
	2	3	5
2 %	0.898	0.910	0.911
5 %	0.897	**0.909**	0.912
10 %	0.892	0.901	0.904
20 %	0.864	0.868	0.868

(b) BDTG (Fakes)

Table 7.8: AUC values of the *BDT (Fakes)* and *BDTG (Fakes)* classifiers evaluated on the total background for different hyperparameter setups. The values get worse with decreasing tree depth and increasing minimum node size. The result for the finally chosen setup is highlighted in bold font.

Classifier	Evaluation on		
	Monte Carlo only	Fakes only	Total background
BDT (MC)	0.782	-	0.798
BDTG (MC)	0.778	-	0.803
BDT (Fakes)	-	0.897	0.897
BDTG (Fakes)	-	0.909	0.909

Table 7.9: AUC values of the trained BDT classifiers evaluated on the different backgrounds. Those of the multijet BDT outperform the ones of the Monte Carlo BDT by far.

7.3 Combination of Tau ID and BDT Classifier Output

In the previous section, two boosted decision trees have been trained independently for the discrimination of the $ZZ \rightarrow 4\tau$ signal against the Monte Carlo and QCD multijet background. For the training, only kinematic variables have been used so that the classifier outputs can now be combined with the tau probability product in order to set up different phase space regions for a fit of the $ZZ \rightarrow 4\tau$ cross section. Additionally, the $N_{\text{b-jets}}$ observable is used for that purpose, whose capability was already motivated in section 7.1. The idea is now to define the fit regions by binning in $N_{\text{b-jets}}$ and both BDT classifiers and perform the fit in the $\log\left(\prod \tau^{\text{prob}}_{\text{i=1-3,4}}\right)$ variable. Overall, there are three control regions (CR), two signal regions (SR) and one validation region (VR) constructed which are depicted in figure 7.11. The corresponding selection criteria for events in these regions can be found in table 7.10. Except for one control region controlling the top-enriched background, a b-veto requirement is placed on all regions. For the *BDTG (MC)* classifier, only a very soft splitting value of -0.8 is chosen since this already rejects most of the background with at least three truth-matched τ-leptons and also a considerable fraction of the fake tau events stemming from heavy particle decays (cf. fig. 7.7). In contrast to that, the *BDTG (Fakes)* distribution is divided into three parts with splitting values at 0.3 and 0.8. The "low" region of *BDTG (Fakes)* < 0.3 contains the majority of QCD multijet events, whereas the signal regions are further split into one with a classifier value between 0.3 and 0.8 and one with a value higher than 0.8. Additionally, a validation region is constructed which needs to fulfill the requirement of *BDTG (Fakes)* > 0.3 as well but of course differs from the signal regions in its *BDTG (MC)* classifier values.
The resulting background composition of the single fit regions is depicted in 7.12. For a better visualization in the pre-fit plots, the fakes are further split into the fraction produced by QCD multijet processes (gray) and the other part stemming from heavy particle decays predicted with the data driven method (orange). Although this procedure requires a different implementation of the τ-fake estimation, it should be noted that it does not change the overall number of expected background events. From the top left plot (a) it can be found that the used BDT classifiers perform quite well with respect to the background they have been trained on. The overwhelming majority of the QCD fakes is separated into the respective control regions, although also a small fraction contaminates the signal regions. The b-quark enriched CR hardly contains any QCD events, which is important in order to control the dominant Monte Carlo background processes $t\bar{t}$ and single top production. In the signal regions (cf. (b)), ratios in $S/\sqrt{B}$ of ≈ 0.02 and ≈ 0.05 can be reached which is already a good result if one compares the number of signal and background events to the preselection level where the background yields were many orders of magnitude higher than those of the signal sample. Subfigures (c) and (d) show the data-to-model agreement in the different fit regions for all events being summed up to one bin. The modeling for the control regions looks very good, whereas in the signal regions the number of background events seems to be slightly underestimated. The single regions are further investigated with respect to their $\log\left(\prod \tau^{\text{prob}}_{\text{i=1-3,4}}\right)$ distributions in the following subsections.

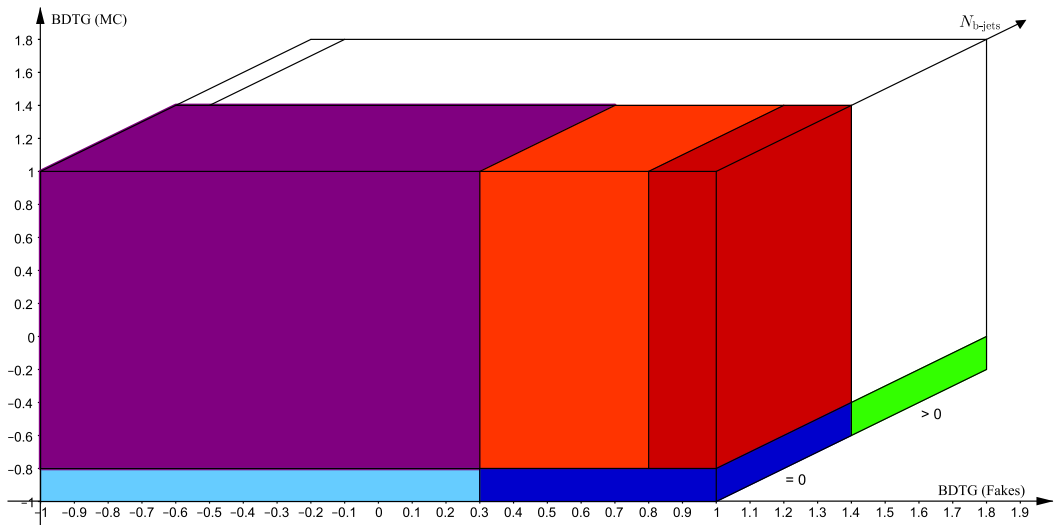


Figure 7.11: The fit regions of the profile likelihood fit for the measurement of the $ZZ \rightarrow 4\tau$ cross section are shown. They are constructed by binning the events in $N_{\text{b-jets}}$ and the *BDTG (MC)* and *BDTG (Fakes)* classifier outputs. At total, there are three control regions (light blue & violet enriched with tau fakes, green enriched with $t\bar{t}$ events), two signal regions (orange & red) and one validation region (dark blue) which is not considered in the fit. Created with [69].

Label	Short label	Type	Selection		
			$N_{\text{b-jets}}$	BDTG (MC)	BDTG (Fakes)
bVeto, MCLow, FLow	$\text{CR}_{\text{MCLowFLow}}$	CR	$= 0$	< -0.8	< 0.3
bVeto, MCHigh, FLow	$\text{CR}_{\text{MCHighFLow}}$	CR	$= 0$	≥ -0.8	< 0.3
bEnr, MCLow, Fgm09	CR_{Top}	CR	> 0	< -0.8	> -0.9
bVeto, MCHigh, FMed	$\text{SR}_{\text{MCHighFMed}}$	SR	$= 0$	≥ -0.8	≥ 0.3 & < 0.8
bVeto, MCHigh, FHigh	$\text{SR}_{\text{MCHighFHigh}}$	SR	$= 0$	≥ -0.8	≥ 0.8
bVeto, MCLow, FMed+	$\text{VR}_{\text{MCLowFMed+}}$	VR	$= 0$	< -0.8	≥ 0.3

Table 7.10: Labeling and selection criteria of the fit regions depicted in figure 7.11. The label color is equal (or similar) to the color of the corresponding region in the illustration. In $N_{\text{b-jets}}$ and *BDTG (MC)* a split in two parts each is done, while the *BDTG (Fakes)* classifier is, at least partially, divided into three phase space regions.

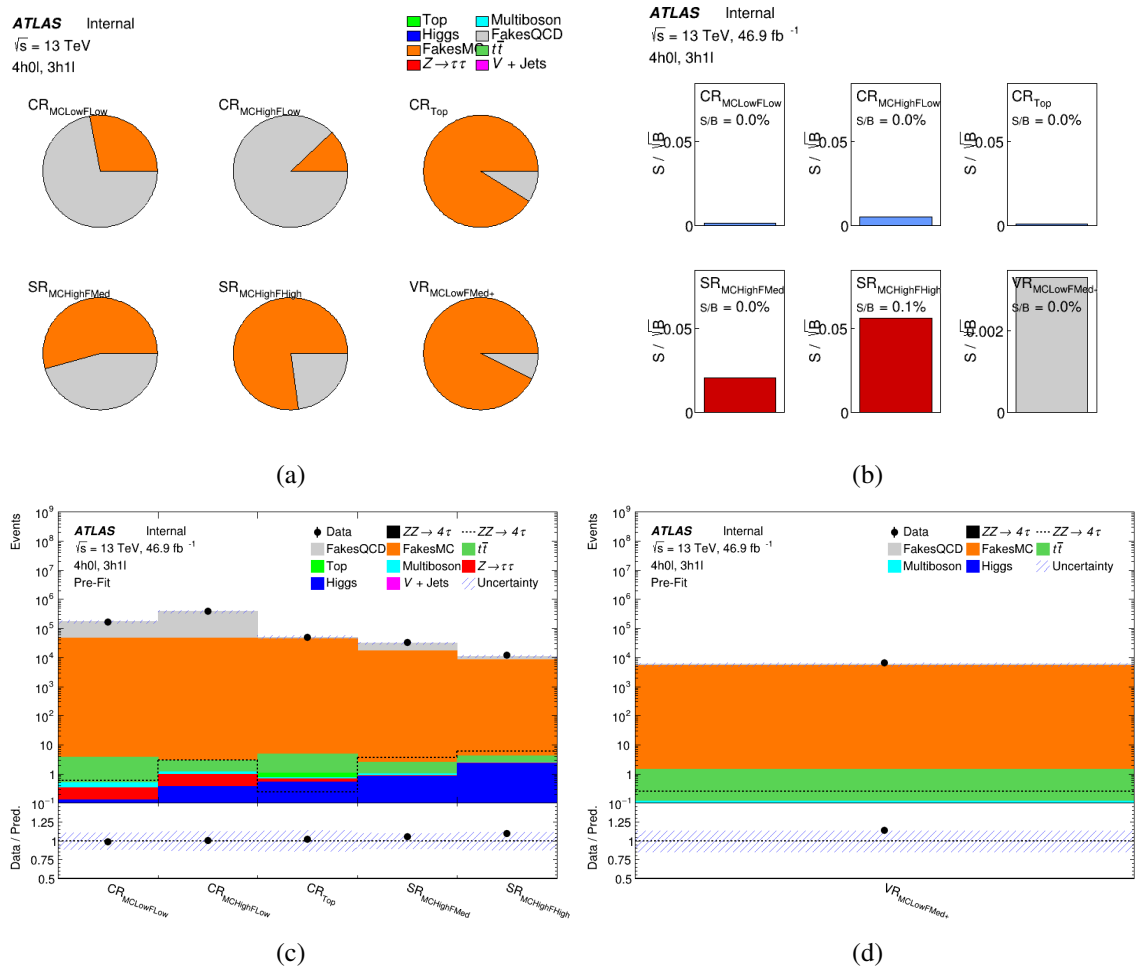
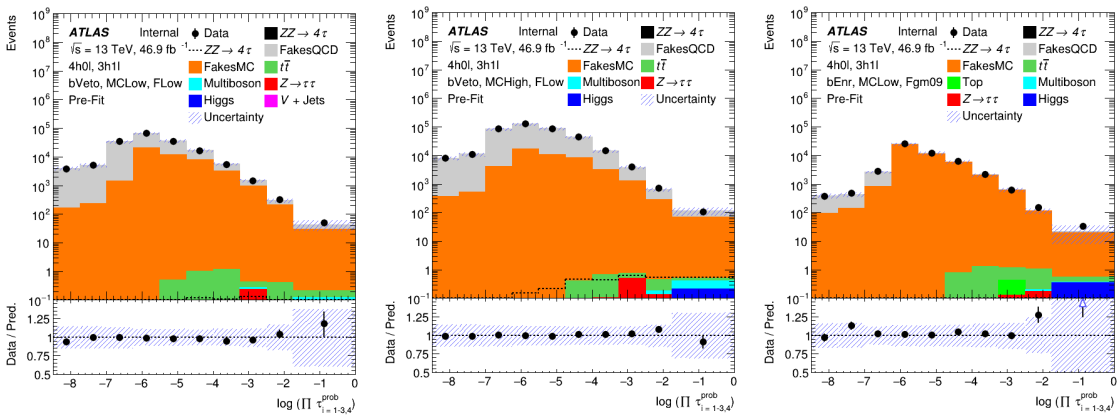


Figure 7.12: In (a), the background composition of the different fit regions is shown. In contrast to the plots in the previous sections, the data driven estimated fakes are now divided into their fraction originating from QCD multijet processes (gray) and the Monte Carlo events with faked τ-leptons (orange). The background contribution from the remaining samples is very low and therefore not visible here. As desired, the multijet fakes mainly accumulate in the regions with a low *BDTG (Fakes)* value. Additionally, (b) displays the signal fraction $S/\sqrt{B}$ in each fit region. While it is close to zero for the control and validations regions, values of ≈ 0.02 and ≈ 0.05 are reached for the signal regions. Finally, the absolute event yields as well as the agreement between the estimated background and the recorded data for the different regions can be seen in (c) and (d). It is found to be very good, especially in the three control regions.

7.3.1 Control Regions

As their name implies, control regions are intended to control the shape and event yields of the different backgrounds relevant for the chosen analysis channel. They are usually constructed for each background and should contain as less events originating from other physical processes as possible. A separate norm factor is then determined for the events in each CR and propagated to the signal regions afterwards. However, in a combined maximum likelihood fit, all bins in all regions are considered simultaneously, so that the notation "control region" only refers to a background-enriched region without a significant signal contribution.
According to the used data driven estimation technique introduced in section 6.6, the background in this setup has only two main important contributions which dominate the expected event yields by far. Therefore, there are just three CRs constructed, from which two ($\mathrm{CR}_{\mathrm{MCLowFLow}}$, $\mathrm{CR}_{\mathrm{MCHighFLow}}$) are used to control the agreement between QCD multijet fakes and data while the third one ($\mathrm{CR}_{\mathrm{Top}}$) serves as a control region for the large $t\bar{t}$ fraction in the remaining, also data driven estimated background. Their resulting distributions in the $\log\left(\prod \tau^{\mathrm{prob}}_{\mathrm{i=1\text{-}3,4}}\right)$ observable are depicted in figure 7.13.

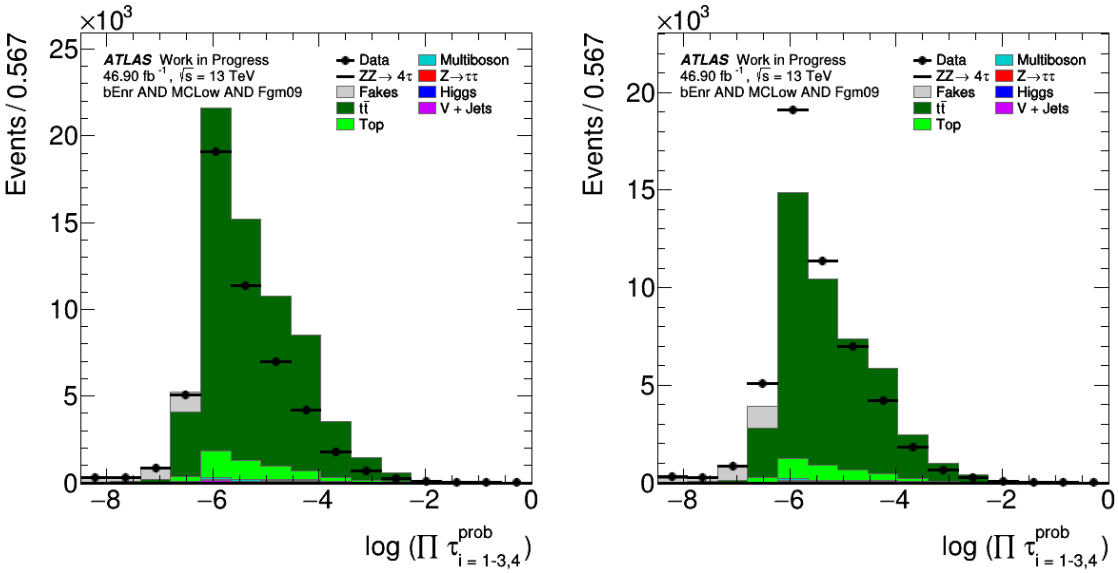


(a) $\mathrm{CR}_{\mathrm{MCLowFLow}}$ (b) $\mathrm{CR}_{\mathrm{MCHighFLow}}$ (c) $\mathrm{CR}_{\mathrm{Top}}$

Figure 7.13: The distribution of the $\log\left(\prod \tau^{\mathrm{prob}}_{\mathrm{i=1\text{-}3,4}}\right)$ observable is presented for the three constructed control regions. In all of them, the data-to-model agreement is found to be very good, only the last bins show some larger deviations. Nevertheless, they are still within the uncertainty band which also increases due to the low number of events in this phase space region.

A good agreement of data and model can be observed over a wide variable range. For $\log\left(\prod \tau^{\mathrm{prob}}_{\mathrm{i=1\text{-}3,4}}\right)$ values close to zero, the statistics are very limited and there are some deviations between the expected and measured event yields, but still within the uncertainty band. The shape of all three regions looks quite similar, only the fraction of the QCD fakes in $\mathrm{CR}_{\mathrm{Top}}$ is significantly smaller than in the other two regions. Furthermore, the gray and orange distributions have their maximum value in the left half of the $\log\left(\prod \tau^{\mathrm{prob}}_{\mathrm{i=1\text{-}3,4}}\right)$ range, while the events with at least three truth-matched τ-leptons visualized in the other colors tend to have higher $\log\left(\prod \tau^{\mathrm{prob}}_{\mathrm{i=1\text{-}3,4}}\right)$ values. This also confirms the expectations and matches the results from e.g. figure 7.1(b).
The actual background composition of the orange histogram in the $\mathrm{CR}_{\mathrm{Top}}$ region can only be checked

qualitatively. For that purpose, the QCD fake estimation procedure presented in section 6.5 is used again, where additionally a truth-matched veto is placed on the background events taken from Monte Carlo samples. Afterwards, the obtained distribution is scaled to the total number of events in the CR_{Top} region. Figure 7.14 shows the result of this procedure together with the distribution where no rescaling is done. Most importantly, the plots reveal that the background is extremely dominated by $t\bar{t}$ and single top events which is desired in order to be able to estimate a reliable scale factor for this part of the background. It also shows a difference between the cases when the rescaling to the CR_{Top} region event yield is applied and when it is not. In the first case, an excess of the predicted background over data is observed whereas in the second plot the agreement is better except for the bins close to the maximum of the distribution. However, this visualization can only be interpreted qualitatively and the deviations do not play any role for the further analysis.



(a) No rescaling

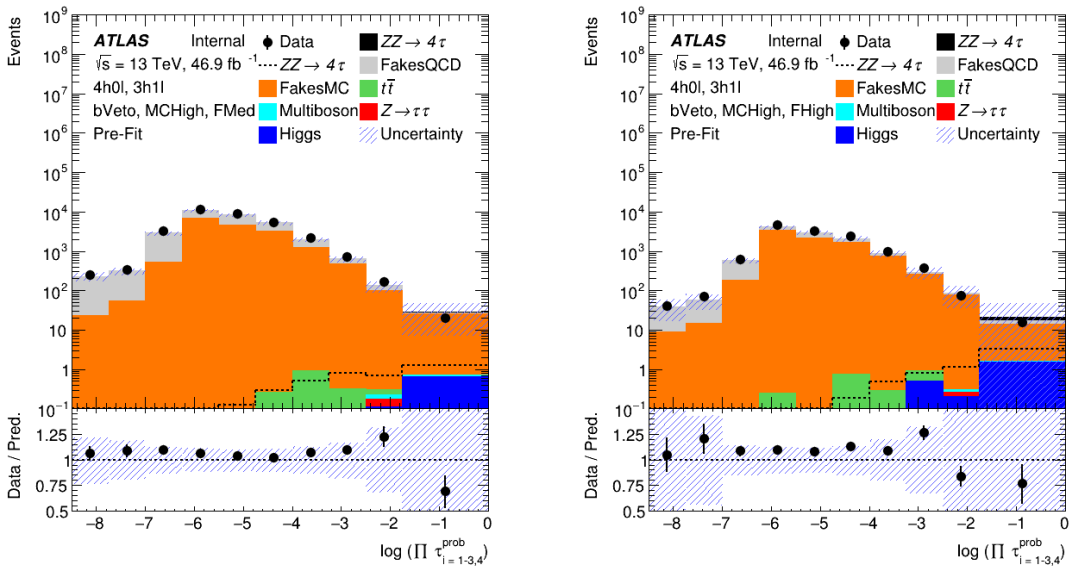
(b) Rescaling included

Figure 7.14: Qualitative pre-fit background composition in the CR_{Top} region constructed with the fake estimation method presented in 6.5 and the application of an additional truth-matched veto on the events taken from Monte Carlo simulation. In the left plot, an excess of the predicted background over the recorded data in nearly the whole variable range can be observed. Hence, the right plot might serve as a more closer approximation as the distribution is rescaled to the total number of events in the CR_{Top} region. Nevertheless, the $t\bar{t}$ and single top purity in the region seems to be close to 100 %.

7.3.2 Signal Regions

In contrast to the CRs, the signal regions should contain the majority of $ZZ \rightarrow 4\tau$ events in order to allow for a precise determination of the observed signal strength compared to the Standard Model expectation. In this analysis, two signal regions are defined which differ in their requirement in the *BDTG (Fakes)* variable. Hence, one region consists of $ZZ \rightarrow 4\tau$ events which have a very high *BDTG (Fakes)* value, while the other one contains events with only a medium classifier score (cf. tab. 7.10). The result is depicted in figure 7.15. Although the agreement between the predicted background and recorded data is not as good as for the control regions presented in the previous section, it is still found

to be within the given statistical uncertainty. The shape of the two dominant backgrounds is very similar to the one in the CRs, however the absolute number of background events is much lower. In the last bin of each plot, the expected signal contribution can be seen on top of the stacked background for the first time. This demonstrates the power of the developed event selection, remembering the expected background on preselection level which was many orders of magnitude higher than the number of signal events.



(a) $SR_{MCHighFMed}$ (b) $SR_{MCHighFHigh}$

Figure 7.15: The distribution of the $\log\left(\prod \tau^{\text{prob}}_{\text{i=1-3,4}}\right)$ observable is presented for the two constructed signal regions. Compared to the control regions, the data-to-model agreement is slightly worse but still within the uncertainty band for most of the bins. The events with faked τ-leptons again accumulate at lower $\log\left(\prod \tau^{\text{prob}}_{\text{i=1-3,4}}\right)$ value, whereas in the last bin of each plot a visible signal contribution can be seen.

7.3.3 Validation Region

Finally, the fit region setup is completed with one validation region. Such regions need to be orthogonal to the other fit regions and are used to determine the accuracy of the background scale factors obtained from fitting the data in the control regions. Thus, the VRs are not fitted themselves, but the results from the fit in the other regions is propagated to them. Figure 7.16 shows the distribution of the $\log\left(\prod \tau^{\text{prob}}_{\text{i=1-3,4}}\right)$ observable in the validation region defined according to table 7.10. From all of these regions, the VR yields the worst data-to-model agreement. Especially in the first and last two bins, the mismodeling becomes visible which also the reason for this regions to be not included in the fit performed in chapter 8. In addition to the pre-fit plots presented in the last subsections, detailed event yield tables for the single fit regions can be found in appendix E.

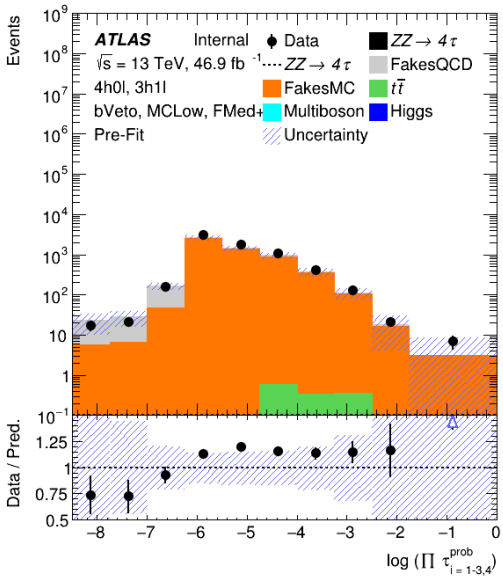


Figure 7.16: The distribution of the $\log\left(\prod \tau^{\text{prob}}_{\text{i=1-3,4}}\right)$ observable is presented for the constructed validation region. Significant deviations between expected background and recorded data can be seen. Furthermore, the uncertainty band is quite large in the first and last bins.

CHAPTER 8

Measuring the $ZZ \rightarrow 4\tau$ Cross Section

In the last chapter, two forests of boosted decision trees were trained and evaluated to enhance the separation of $ZZ \rightarrow 4\tau$ signal events from the expected background. The obtained multivariate classifiers have been used to construct different phase space regions which are either enriched with signal events or not. These regions then can be taken as input for a profile likelihood (PLH) fit which finally yields a result for the observed $ZZ \rightarrow 4\tau$ cross section compared to the Standard Model prediction. For that purpose, `TRExFitter` [70] is used which is an ATLAS internal framework for binned template profile likelihood fits. The method of a PLH fit is briefly introduced in section 8.1. Afterwards, section 8.2 gives an overview of the systematic uncertainties included in this analysis, before the fit results are presented in section 8.3. The chapter is concluded with section 8.4, where the obtained results are interpreted in the context of the Standard Model.

8.1 Profile Likelihood Fit Procedure

In most high energy particle physics analyses, the experimenters aim to measure one or more certain physical quantities, referred to as parameters of interest (POI). This analysis has only one POI, which is the $ZZ \rightarrow 4\tau$ signal strength $\mu(ZZ \rightarrow 4\tau)$. It is defined such that $\mu = 1$ represents the Standard Model expectation, $\mu = 2$ twice the expectation and so on, while a value of $\mu = 0$ means that the process does not occur at all. Now, one wants to find an estimator for this parameter of interest which approximates its true value. A common way to do that is to use the so-called maximum likelihood method [71]. Assuming that the signal and background predictions in each bin of a phase space region depend on parameters (μ, θ), a binned global likelihood function can be constructed in the following way [72]:

$$\mathcal{L}(n|\mu, \theta) = \prod_{i \in \text{bins}} \mathcal{P}\left(n_i | \mu \cdot S_i(\theta) + B_i(\theta)\right) \quad . \tag{8.1}$$

The number of observed events in the bin i is denoted with n_i, whereas $\mathcal{P}$ states the Poisson distribution of the signal plus background prediction $S + B$ in the bin i. The desired result is then simply the value of the POI μ which maximizes the likelihood function. However, beside the pure statistical uncertainty there are usually also systematic uncertainties which take the imperfect knowledge of certain model parameters into account, e.g. the resolution in data and mismodeling resulting from it. Since these systematic effects have a direct impact on the S and B predictions in the single bins, they need to be included in the likelihood as well. The idea is to treat them as "constrained" nuisance

parameters (NP), defined as $\theta = \theta^0 \pm \Delta\theta$ with $\theta^0 = 0$ and $\Delta\theta = 1$ fixed by convention. Hence, the effect of a systematic j on the prediction x in bin i is defined for $\theta = -1$ and $\theta = +1$ and is inter- or extrapolated for any other value of θ. It is then implemented in the likelihood as an additional constraint, i.e. a probability density function (usually a Gaussian $\mathcal{G}$, so that $\pm\theta$ refers to one Gaussian standard deviation), resulting in [72]:

$$\mathcal{L}\left(n, \theta^0 | \mu, \theta\right) = \prod_{i \in \text{bins}} \mathcal{P}\left(n_i | \mu \cdot S_i(\theta) + B_i(\theta)\right) \times \prod_{j \in \text{syst}} \mathcal{G}\left(\theta_j^0 | \theta_j, \Delta\theta_j\right) \quad . \tag{8.2}$$

Beside systematic uncertainties, further free parameters for particular S or B components can be included in the likelihood. They are called normalization (or norm) factors k and act on a component as $B(\theta, k) = k \cdot B(\theta)$. Last, there is an additional nuisance parameter for the statistical uncertainty originating from the limited number of Monte Carlo events used to predict the number of signal and background events in each bin. These NPs in form of Poisson priors are independent from each other and are named "gammas" γ. Having the likelihood for one phase space region defined, one can now take the likelihood product of the single regions to obtain the function describing the whole analysis space, as long as the regions are orthogonal, i.e. do not overlap in the events they contain. The determination of the estimators for the POI and all nuisance parameters is then done by carrying out a N-dimensional likelihood maximization (or negative-log likelihood minimization). If the best-fit values for the norm factors differ from their pre-fit assumption, the background prediction in the fit regions might change compared to the pre-fit state. Furthermore, the fit can constrain single nuisance parameters from the statistical power of the data, resulting in a reduced uncertainty on the background expectation [72], or pull[1] them.

8.2 Systematic Uncertainties

One of the goals of the analysis performed in this thesis is to validate the multi-tau event identification method developed in [8] and described in section 5.2 on a Standard Model process. Therefore, especially the impact of tau efficiency related scale factors[2] and systematic uncertainties on the multiplied probability method is of interest. However, this treatment is not trivial and is therefore briefly explained in the following.

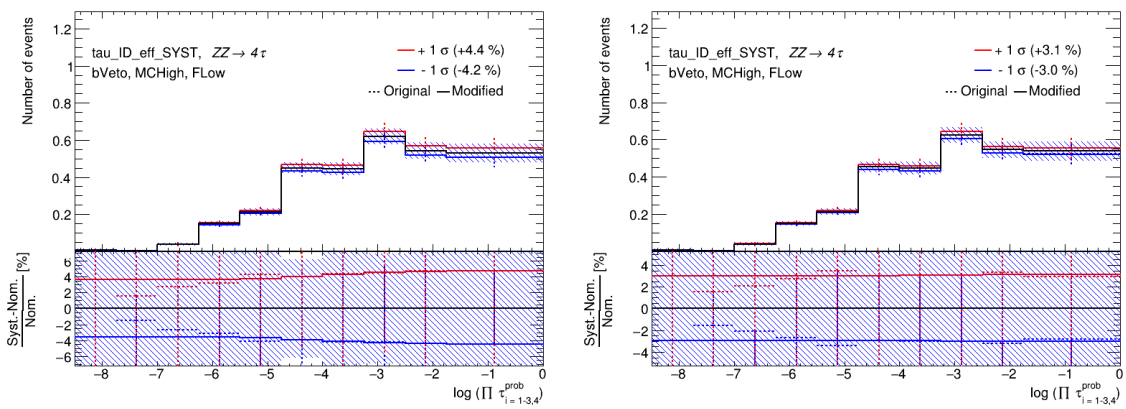
8.2.1 Tau ID Uncertainties

As introduced in section 5.1, the "classic" tau identification approach used in ATLAS requires each tau candidate to fulfill one of the working points *very loose*, *loose*, *medium* or *tight*. Therefore, the systematic uncertainties related to the tau ID also are only defined for these fixed inclusive working points and using them with the identification method of multiplied probabilities is technically not possible correctly. In order to tackle this problem, two options have been tested. The first possibility is to use the systematic of the *loose* working point for all events and truth-matched tau candidates, as the majority of them is not expected to fulfill more than exactly this working point in terms of the τ^{jetBDT} score. The second option would be to evaluate for event by event which working point the tau

[1] A pull corresponds to a deviation from the prior pre-fit value of the NP after the fit.

[2] These scale factors correspond to an efficiency correction between Monte Carlo simulation and data which is applied to each τ-lepton and then changes by a certain factor for each systematic variation.

candidate would hypothetically fulfill and then use the corresponding systematic uncertainty. Figure 8.1 exemplary compares the resulting distributions of the *tau_ID_eff_SYST* systematic containing different detector effects which need to be taken into account (see later in this section) for the $\mathrm{CR}_{\mathrm{MCHighFLow}}$ region. Neither in their shape nor in their absolute normalization a significant difference is visible. However, since it is more intuitive to use for each tau candidate and event the uncertainty related to the working point the tau fulfills in terms of its τ^{jetBDT} score, the second of the discussed options is chosen for the further fit procedure.
In this analysis, p_{T}-dependent systematic uncertainties on the recurrent neural network (RNN) based offline tau identification efficiency and the tau trigger identification efficiency are considered. Furthermore, a prong-dependent uncertainty on the tau energy scale as well as a η-dependent uncertainty on the correction factor for the probability of misidentifying an electron as a hadronic τ-lepton decay are taken into account. These values are usually measured by tag-and-probe analyses in the $Z \rightarrow \tau_{\mu}\tau_{\mathrm{had}}$ and $Z \rightarrow ee$ decay channels (see [73] for more detailed information), their current values can be found in [74] and [75]. Additionally, the *tau_ID_eff_SYST* systematic mentioned already above includes different detector effects which need to be treated correlated, especially uncertainties concerning jet, muon and MET reconstruction and identification.



(a) Fixed *loose* tau working point

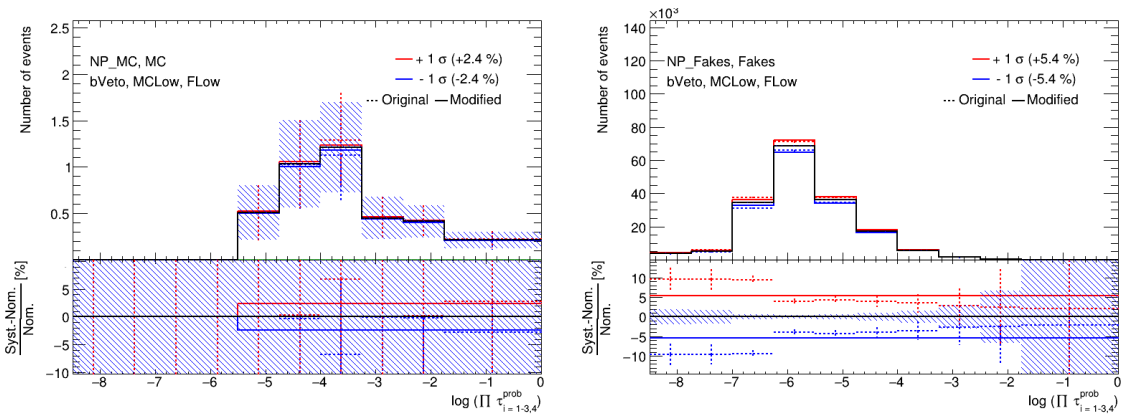
(b) Variable tau working point

Figure 8.1: Distribution of the *tau_ID_eff_SYST* systematic uncertainty in the $\mathrm{CR}_{\mathrm{MCHighFLow}}$ region for the two different approaches of using them with the multiplied probability tau identification method. In the shape hardly any difference can be observed and in their normalization they also differ only on the level of approximately 1 %.

8.2.2 Further Uncertainties

Beside the tau systematics discussed in the previous section, no further instrumental uncertainties are included at the current state of the analysis. However, there is one more systematic representing the uncertainty in the data driven fake tau estimation considered and also the scale factor for the few remaining background events taken from Monte Carlo simulations is not implemented as a free norm factor in the fit, but as systematic uncertainty instead. Therefore, each of the two systematics is defined with an overall up and down shift. For the one acting on the Monte Carlo background, named *NP_MC*, the variation is chosen to be 100 % since their absolute number compared to the data driven estimated events is very small and also a variation of +1 or −1 would not lead to a significant change

in the total number of estimated background events. This gives the fit the possibility to constrain the uncertainty itself in the best possible way. A reasonable prior variation for the fake tau estimation systematic (*NP_Fakes*) can be motivated from table 6.3. For data, the expectation from the binomial distribution for the ratio of events in the unphysical charge regions is 1 : 4 : 6 : 4 : 1. The observed ratio deviates from that as 1.0 : 4.4 : 6.6 : 4.1 : 0.9, corresponding to a value of approximately 10 %. Since it is more likely that this deviation is underestimated than that it is overestimated, another 5 % are added to the prior which is finally set to 15 % for as well the up as the down variation. The resulting distributions of these two nuisance parameters is exemplary shown for the $\mathrm{CR}_{\mathrm{MCLowFLow}}$ region in figure 8.2. Albeit the prior shift was chosen quite large, the visible variations are only in the order of a few percent. The most noticeable difference for both nuisance parameters is the width of their statistical error band which is way larger for *NP_MC* than for *NP_Fakes* and originates from the very low number of events in the remaining Monte Carlo background on which the *NP_MC* systematic acts. After the application of a smoothing procedure, the ratio of the deviation from nominal and nominal itself is nearly constant for both systematics.



(a) NP_MC (b) NP_Fakes

Figure 8.2: Distribution of the *NP_MC* and *NP_Fakes* nuisance parameters in the $\mathrm{CR}_{\mathrm{MCLowFLow}}$ region. Both do not show any conspicuous behavior, although the statistical error band in *NP_MC* is much broader than for *NP_Fakes*, caused by the very low number of events. The effect of the smoothing procedure (dashed line before, solid line after) can be nicely seen in the ratio plot at the bottom of each figure.

8.2.3 Pruning

Usually, there is a large number of systematic uncertainties which need to be included in the likelihood function, increasing the complexity and also the computing time of the performed fit. Therefore, it is useful to check the impact of these nuisance parameters before to make the fit procedure more efficient. In this setup, a comparably small value of 0.1 % has been chosen for the lower threshold defining when a nuisance parameter is dropped, for both shape and normalization systematics. The result of the pruning process for the given systematics introduced in the last subsections is depicted in figure 8.3. It can be seen that only for the signal sample a few nuisance parameters are kept. In all other cases, the corresponding effects are too small so that either the shape of the systematic is dropped and only the absolute normalization is kept, or it is dropped completely. Hence, the uncertainties on

the tau identification efficiencies and other related effects do not have a large impact on the method of multiplied probabilities. After this last check and performance optimization step, finally the fit can be carried out by minimizing the negative-log likelihood function (using the *Minuit2* minimizer implemented in ROOT [56]). The results are shown in the next section.

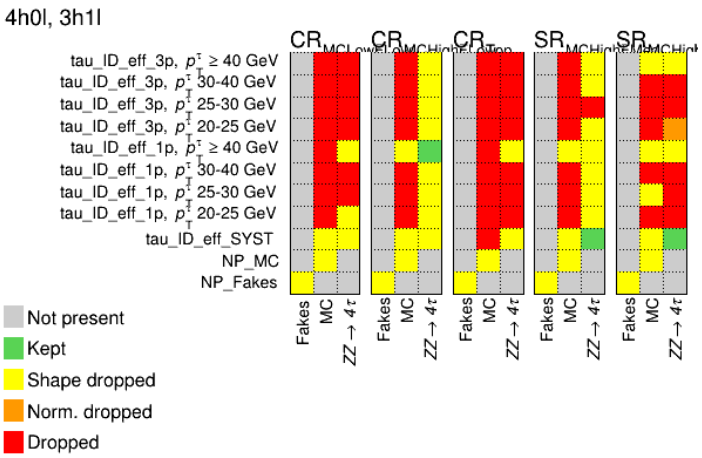


Figure 8.3: Result of the pruning applied to all defined nuisance parameters. The tau ID systematics are only applied to samples with truth-matched τ-leptons, i.e. to all except for the data driven estimated fakes. The nuisance parameters *NP_MC* and *NP_Fakes* act only on the respective sample. Only a few systematics survive the pruning process, in most of the case either their shape is dropped and only the normalization is kept or they are dropped completely. This is especially true for the p_T-dependent uncertainty on the tau identification efficiency. The different regions read from left to right: $CR_{MCLowFLow}$, $CR_{MCHighFLow}$, CR_{Top}, $SR_{MCHighFMed}$ and $SR_{MCHighFHigh}$.

8.3 Fit Results

8.3.1 Asimov Data

In order to test the technical functionality of the chosen setup, first a $S + B$ fit over all defined control and signal regions on Asimov data is performed. Such an Asimov dataset is constructed from the respective sum of predicted signal and background events in each bin. Consequently, one expects the fit results for the POI and all nuisance parameters to be equal to their nominal value, i.e. one for the $ZZ \to 4\tau$ signal strength and zero for each NP. If the obtained results deviate from these values, this might indicate that something is not working properly. Furthermore, the fit on Asimov data can already give some hints about the expected order of magnitude of the errors on the POI and the nuisance parameters, the correlations between them and possible constraints on certain NPs. While in the pre-fit plots shown in section 7.3 the processes contributing to the fakes and Monte Carlo background samples were visualized separately, they are now treated together again for the fit. In the fakes sample the QCD multijet events as well as those from heavy particle decays are contained, both at least once faking the signature of a hadronically decaying τ-lepton. The remaining underlying events with only truth-matched taus are merged into a combined MC background sample. Figure 8.4 shows the values of $\mu(ZZ \to 4\tau)$ and all nuisance parameters as well as the correlations between

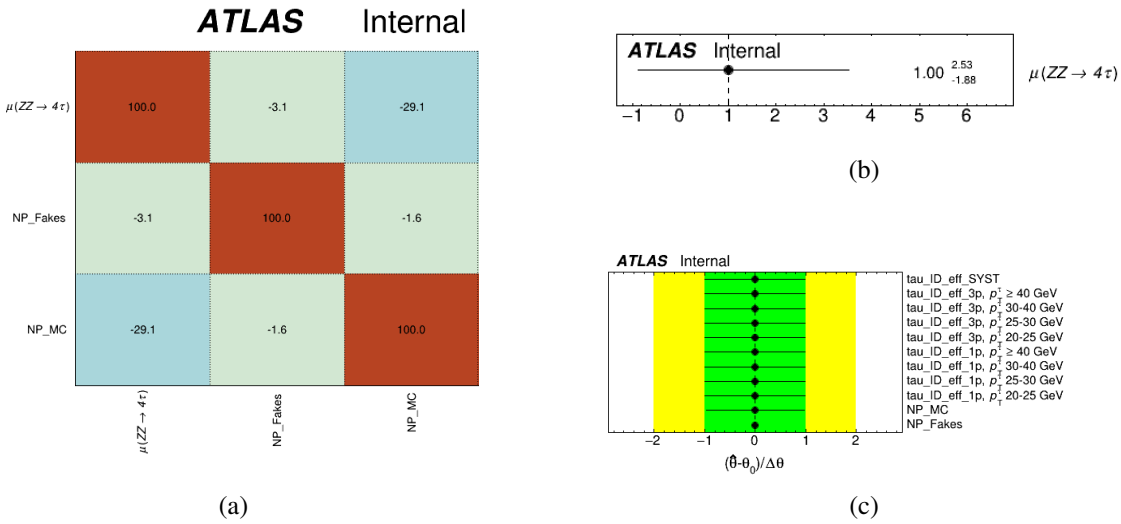
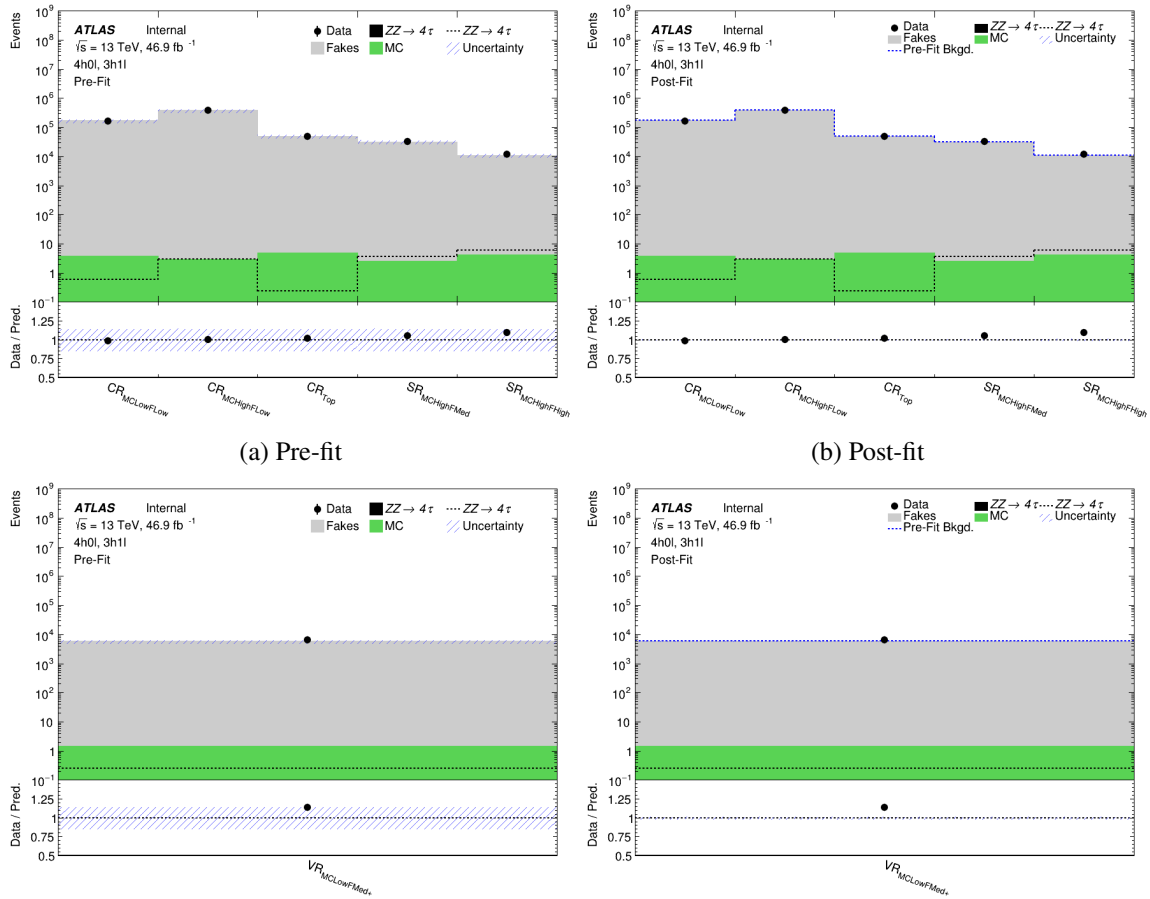


Figure 8.4: The results from the fit on Asimov data are shown. In (a), the correlations between the POI and the nuisance parameters exceeding the set threshold of 10 % are visible. A small correlation between $\mu(ZZ \rightarrow 4\tau)$ and *NP_MC* can be observed. Subfigures (b) and (c) depict the best fit result for the POI and all nuisance parameters, respectively. As expected, always the nominal value is returned. The upper and lower errors on the POI are approximately 200 %, but the upper one is larger than the lower one. The plot of the nuisance parameters indicates that only *NP_Fakes* can be constrained very well.

them resulting from the fit. The POI is measured to be $1.00^{+2.53}_{-1.88}$. This result matches the expectation for the case when the data in each bin is exactly the sum of predicted signal and background events. The error on $\mu(ZZ \rightarrow 4\tau)$ is relatively large and is assumed to mainly originate from the low number of expected signal events and the in general low statistics in the signal-enriched bins in both SRs (cf. fig. 7.15 again). The best fit result for each nuisance parameter is its nominal value, i.e. zero, too. Nevertheless, it can already be observed that, in contrast to all other parameters, the *NP_Fakes* uncertainty can be constrained very well from the large amount of fake tau background events. Hence, it is also hardly correlated to any other nuisance parameter. However, there is a small correlation between $\mu(ZZ \rightarrow 4\tau)$ and *NP_MC*. This can be explained as follows. As the NP for the dominating fakes is constrained very well, and there are only two more parameters which can be considerably varied in the fit, each change in one of them immediately results in a change of the second one in the opposite direction as well. The effect of this constraint can additionally be seen in figure 8.5 where the summary plots of the pre-fit and post-fit state for the single fit regions are compared. Albeit the fit does not cause any modification in the expected background and signal events, it can though significantly reduce the total uncertainty not only in the control and signal regions, but also in the validation region which did not enter the fit. From the obtained results the expected upper limit on $\mu(ZZ \rightarrow 4\tau)$ at a confidence level of 95 % can be calculated using the CL_s method [76][77]. Table 8.1 summarizes the outcome of this calculation. The $ZZ \rightarrow 4\tau$ process is expected to be excluded above a signal strength of 5.162 times the Standard Model expectation at a confidence level of 95 %. In the next section, this fit procedure is repeated using real data recorded by the ATLAS detector.



(a) Pre-fit (b) Post-fit

(c) Pre-fit (d) Post-fit

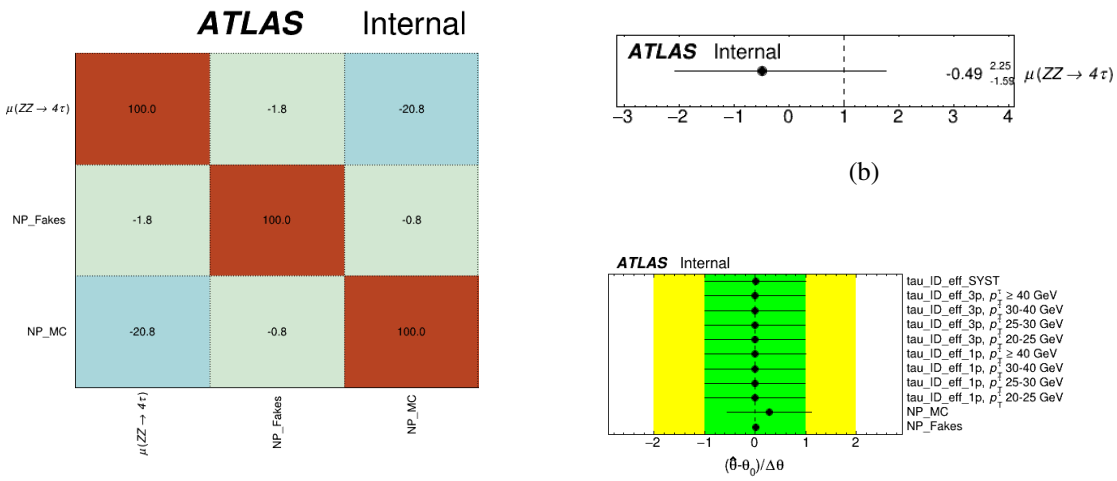
Figure 8.5: The data-to-model agreement in the different fit regions between the pre-fit and post-fit state is compared for the fit on Asimov data. As expected, no difference in the upper part of each plot is observed. Nevertheless, the fit can reduce the total uncertainty in the ratio of Asimov data and predicted background.

Exp. limit	Exp. limit - 1σ	Exp. limit + 1σ	Exp. limit - 2σ	Exp. limit + 2σ
5.162	3.719	8.119	2.77	12.73

Table 8.1: The expected upper limit on $\mu(ZZ \rightarrow 4\tau)$ and the $\pm 1\sigma$ and $\pm 2\sigma$ values calculated from Asimov data are shown. With the currently available statistics, a limit of 5.162 times the Standard Model expectation can be set on the $ZZ \rightarrow 4\tau$ signal strength at a confidence level of 95 %.

8.3.2 ATLAS Data

After the results of the fit on Asimov data presented in the last subsection already gave some hints towards possible constraints on single nuisance parameters, the order of magnitude of the error on the POI and the upper limit on $\mu(ZZ \to 4\tau)$ one can expect, the profile likelihood fit is now performed on the data recorded by the ATLAS detector in 2017 with an integrated luminosity of 46.9 fb^{-1}. Figure 8.6 shows the obtained results for the $ZZ \to 4\tau$ signal strength, the nuisance parameters and the correlations between them. The numerical values are additionally given in table 8.2. Different things can be seen. First, the fitted value of $\mu(ZZ \to 4\tau)$ is $-0.49^{+2.25}_{-1.59}$ and thus negative. This result can be explained as follows. Looking back to figure 7.13, the predicted background in the two last bins seems to be underestimated (except for the last bin in the $\mathrm{CR}_{\mathrm{MCHighFLow}}$ region). Since the estimated fake background is fixed extremely well by the remaining bins in all regions, this underestimation in the last bins can only be compensated by scaling up the MC background (i.e. everything except for the gray and orange distributions in figure 7.13). As in the signal regions (cf. fig. 7.15) already a lower number of data events is recorded in the last bins than the background which is predicted, the only possibility to account for the increased MC background is then to scale down the signal distribution accordingly, resulting in a negative signal strength $\mu(ZZ \to 4\tau)$. Second, a significant pull, i.e. deviation from the nominal value, for *NP_MC* can be observed. This exactly matches the behavior which has been described before. Furthermore, very small pulls for *NP_Fakes* and *tau_ID_eff_SYST* can be seen, but their effect is negligible. The third consequence is the non-vanishing correlation between the signal strength and *NP_MC*, occurring due to the same reason as the nuisance parameter pull.



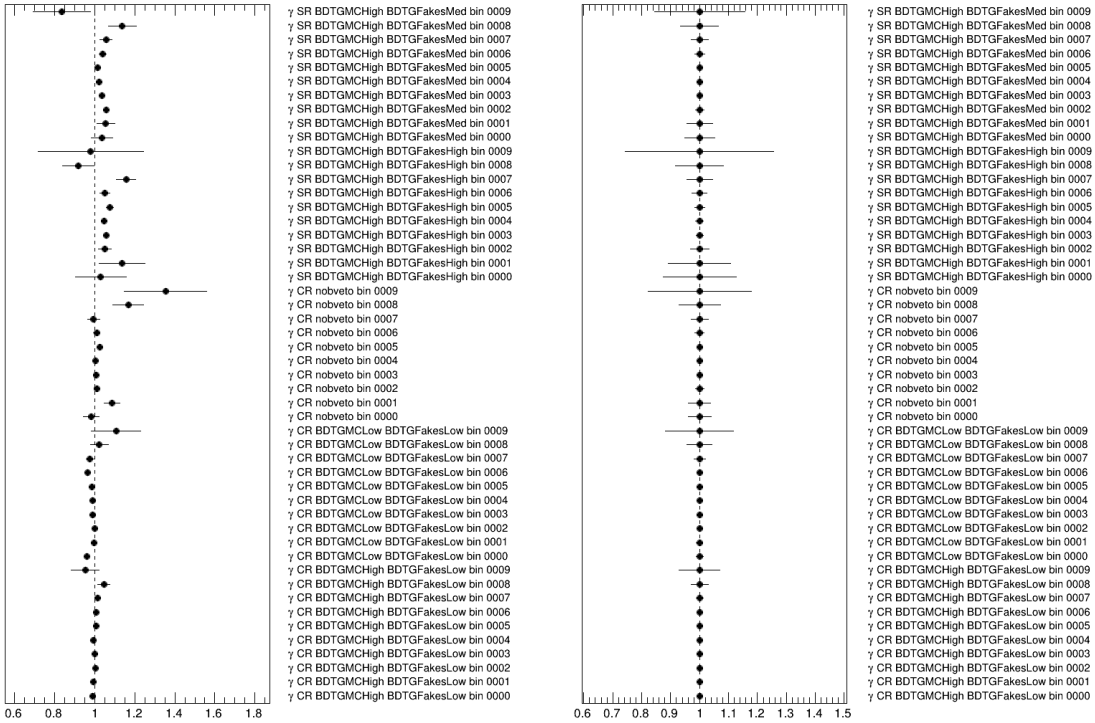
(a) (b) (c)

Figure 8.6: The results from the fit on ATLAS 2017 data with an integrated luminosity of 46.9 fb^{-1} are shown. In (a), the correlations between the POI and the nuisance parameters exceeding the set threshold of 10 % are visible. Again, a small correlation between $\mu(ZZ \to 4\tau)$ and *NP_MC* can be observed. Subfigures (b) depicts the best fit result for the POI. The obtained $ZZ \to 4\tau$ signal strength is negative, but the Standard Model expectation of $\mu = 1$ still lies within the error range which is slightly reduced compared to the Asimov fit. In (c), the resulting values for the nuisance parameters are given. Similar to figure 8.4(c) the *NP_Fakes* parameter can be constrained quite well. Additionally, the plot now indicates a very small pull, i.e. deviation from the nominal value, for *tau_ID_eff_SYST* and *NP_Fakes* as well as a significant one for *NP_MC*.

Parameter	Fit value	Upper error	Lower error
$\mu(ZZ \to 4\tau)$	-0.487	+2.248	-1.593
tau_ID_eff_SYST	0.0013	+0.9934	-0.9934
tau_ID_eff_3p, $p_{\mathrm{T}}^{\tau} \geq 40\,\mathrm{GeV}$	0.0003	+0.9933	-0.9933
tau_ID_eff_3p, p_{T}^{τ} 30-40 GeV	1.27×10^{-6}	+0.9933	-0.9933
tau_ID_eff_3p, p_{T}^{τ} 25-30 GeV	-4.38×10^{-6}	+0.9934	-0.9934
tau_ID_eff_3p, p_{T}^{τ} 20-25 GeV	-3.55×10^{-5}	+0.9933	-0.9933
tau_ID_eff_1p, $p_{\mathrm{T}}^{\tau} \geq 40\,\mathrm{GeV}$	0.0005	+0.9933	-0.9933
tau_ID_eff_1p, p_{T}^{τ} 30-40 GeV	-8.16×10^{-7}	+0.9933	-0.9933
tau_ID_eff_1p, p_{T}^{τ} 25-30 GeV	0.0002	+0.9933	-0.9933
tau_ID_eff_1p, p_{T}^{τ} 20-25 GeV	9.49×10^{-7}	+0.9933	-0.9933
NP_MC	0.278	+0.841	-0.841
NP_Fakes	0.016	+0.013	-0.013

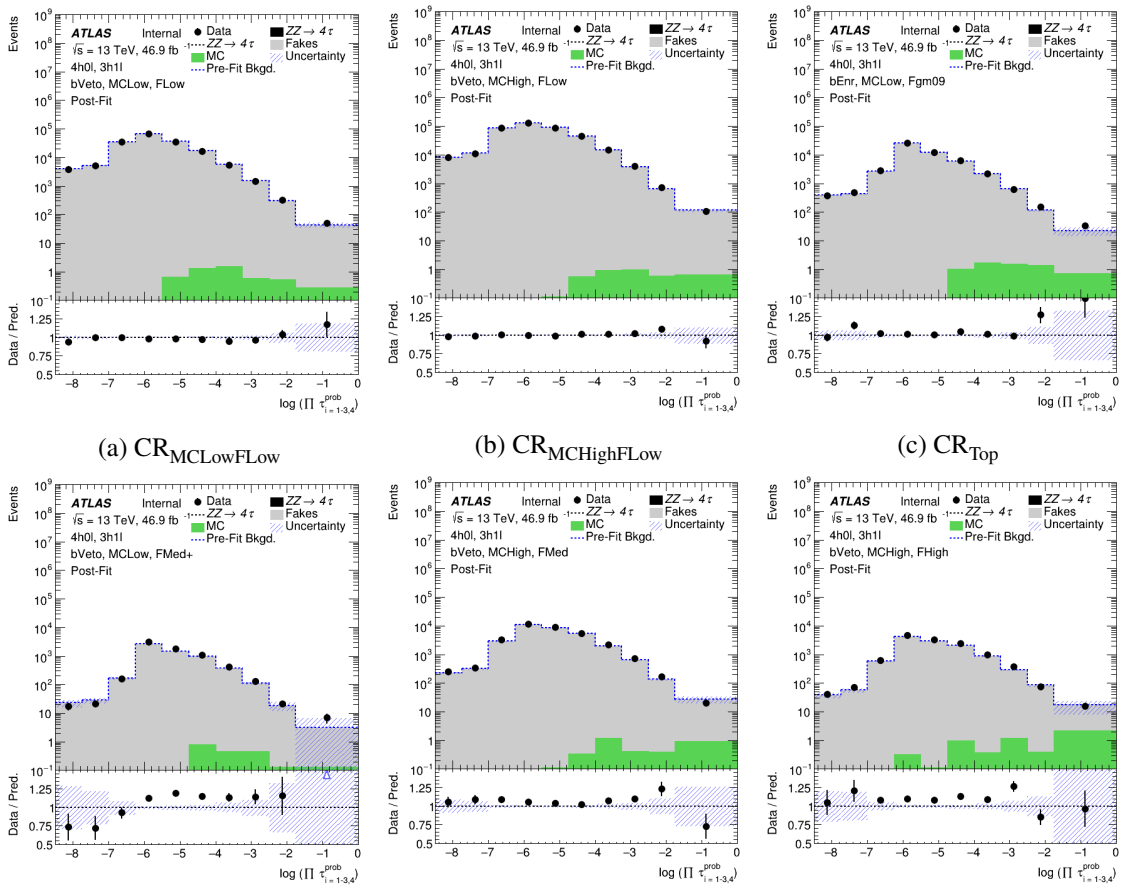
Table 8.2: Resulting post-fit values of the POI and the nuisance parameters using ATLAS 2017 data. Only *NP_Fakes* can be constrained, the errors on all other parameters are quite large. For further explanations see figure 8.6.

In addition to understanding these effects, also the fit results for the gammas, which take the finite Monte Carlo statistics in each bin into account, need to be checked. They are shown in figure 8.7. For comparison, the gammas obtained from the fit on Asimov data are depicted, too. In the left plot it can be seen that most of them are very close to their nominal value of one. Nevertheless, especially for the last bin in the different regions some nuisance parameter pulls can be observed. In these cases, the statistical uncertainty on the gamma is also quite large which is assumed to originate from the low event yields in the respective bins. The gammas obtained from the fit on Asimov data show uncertainties in a similar order of magnitude, albeit the result for the values themselves is always equal to one, as expected when the data in each bin corresponds to the sum of predicted signal and background events. Hence, these results do not show any conspicuous behavior and look consistent. In a next step, post-fit plots of the data-to-model agreement in the different fit regions can be produced. They are given in figure 8.8. In the total background hardly any difference compared to the pre-fit state is visible. But comparing the plots with figures 7.13, 7.15 and 7.16, it can be seen that the signal contribution vanished completely and the number of predicted MC events increased. This behavior reflects the results presented in figure 8.6. It should be noted at this point, that the post-fit effect of the gammas is not propagated to the plots. Furthermore, the total uncertainty in the single bins is reduced significantly through the fit. This applies to the validation region as well which even was not part of the fit at all.



(a) ATLAS data (b) Asimov data

Figure 8.7: The plots show the best fit values for the gammas which take the effect of finite Monte Carlo statistics in the background prediction for the single bins into account. In (a), the result for the fit on ATLAS data is depicted. Except for a few cases, the values are very close to their nominal value of one. Pulls and larger errors mostly occur for the last bin in the different regions. Subfigure (b) states the results of the fit on Asimov data from the previous subsection for comparison. Albeit the gammas all yield their nominal value, the behavior in their errors is similar to the case using real data. The bin numbers in each region are sorted in descending order, i.e. the last bin is given first and the first one last.



(a) $CR_{MCLowFLow}$ (b) $CR_{MCHighFLow}$ (c) CR_{Top}

(d) $VR_{MCLowFMed+}$ (e) $SR_{MCHighFMed}$ (f) $SR_{MCHighFHigh}$

Figure 8.8: The post-fit data-to-model agreement in the different fit regions for the fit on ATLAS 2017 data is shown. Additionally, the pre-fit background sum is visualized as a dashed blue line. Hardly any differences in the total background compared to the pre-fit state can be observed. However, there is no signal contribution in the plots anymore as the best fit value for $\mu(ZZ \rightarrow 4\tau)$ is negative.

Lastly, the impact of the single nuisance parameters on the uncertainty of the POI can be investigated. Therefore, for each NP four fits are performed where the parameter is fixed to the following configurations:

- pre-fit value + pre-fit uncertainty
- pre-fit value − pre-fit uncertainty
- post-fit value + post-fit uncertainty
- post-fit value − post-fit uncertainty .

The "impact" on $\mu(ZZ \rightarrow 4\tau)$ is then calculated as the difference between the signal strength resulting from each of these four fits and $\mu(ZZ \rightarrow 4\tau)$ obtained in the nominal fit (fig. 8.6). The results are visualized in figure 8.9. For none of the tau ID related systematic uncertainties a visible impact on the

value of the POI can be observed. The same applies to *NP_MC*. The $ZZ \rightarrow 4\tau$ signal strength only changes if the *NP_Fakes* parameter is set to its pre-fit value minus the pre-fit uncertainty. In that case it seems that the lower number of background events resulting from this shift is compensated with an increase of signal events. These observations confirm the expectations already made in the pruning step, namely that the systematic effects are dominated by the background estimation procedure while tau identification uncertainties only play a minor role.

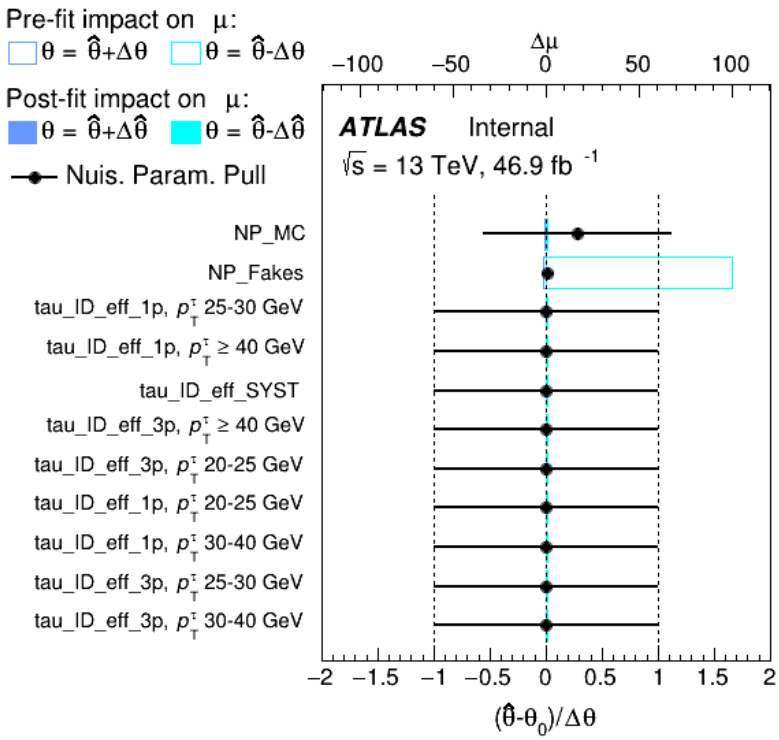


Figure 8.9: The plot shows the impact of the different nuisance parameters on the signal strength $\mu(ZZ \rightarrow 4\tau)$. Only setting the *NP_Fakes* parameter to its pre-fit value minus its uncertainty has a visible effect on the POI.

8.4 Interpretation in the Standard Model

In order to finally interpret the results presented in the previous section in the context of the Standard Model, the expected but also the observed upper limit on the $ZZ \rightarrow 4\tau$ signal strength can be calculated. Therefore, again the CL_s method [76][77] is used. Table 8.3 summarizes the results of this calculation. The signal hypothesis is expected to be excluded for values of $\mu(ZZ \rightarrow 4\tau)$ larger than 5.488 times the Standard Model expectation at a confidence level of 95 %. In contrast to that, the observed upper limit is given by $5.078 \times \mu^{\mathrm{SM}}(ZZ \rightarrow 4\tau)$.
Furthermore, the expected p-value and significance can be obtained. When testing a hypothesis H_1 with the presence of signal (i.e. $\mu(ZZ \rightarrow 4\tau) > 0$) against one H_0 without signal ($\mu(ZZ \rightarrow 4\tau) = 0$), the likelihood ratio $\mathcal{L}(H_0)/\mathcal{L}(H_1)$ is the best possible discriminator. For the profile likelihood method,

Exp. limit	Exp. limit - 1σ	Exp. limit + 1σ	Exp. limit - 2σ	Exp. limit + 2σ	Obs. limit
5.488	3.955	8.762	2.946	13.8	5.078

Table 8.3: The expected upper limit on $\mu(ZZ \to 4\tau)$, the $\pm 1\sigma$ and $\pm 2\sigma$ values and the observed limit calculated from ATLAS 2017 data are shown. With the currently available statistics, a limit of 5.488 times the Standard Model expectation can be set on the $ZZ \to 4\tau$ signal strength at a confidence level of 95 %. The observed limit is $5.078 \times \mu^{\mathrm{SM}}(ZZ \to 4\tau)$.

this ratio is adjusted to be [72]:

$$\lambda(\mu) = \frac{\mathcal{L}(\mu, \hat{\hat{\theta}}_\mu)}{\mathcal{L}(\hat{\mu}, \hat{\theta})} \quad . \tag{8.3}$$

While in the numerator $\mathcal{L}$ is maximized for a given μ (so-called "conditional likelihood"), the denominator denotes the "unconditional likelihood". The test statistics can then be defined as [72]

$$t_\mu = -2 \ln \lambda(\mu) \qquad , \qquad q_0 = \begin{cases} -2 \ln \lambda(0) & \hat{\mu} \geq 0 \\ 0 & \hat{\mu} < 0 \end{cases} \tag{8.4}$$

and p-value p_0 and significance Z_0 are given by [72]:

$$p_0 = \int_{q_{0,\mathrm{obs}}}^{\infty} f(q_0|0)\, \mathrm{d}q_0 \qquad , \qquad Z_0 = \Phi^{-1}(1 - p_0) \ . \tag{8.5}$$

In these equations, Φ^{-1} is the quantile of the standard Gaussian and f the distribution of the test statistics. In data samples with large statistics, this distribution is known according to Wilks' theorem[3] and given by a χ^2 distribution of the form [72]:

$$f = -2 \ln \lambda(\mu) = -2 \left(\ln \mathcal{L}(\mu, \hat{\hat{\theta}}) - \ln \mathcal{L}(\hat{\mu}, \hat{\theta}) \right) = \left(\frac{\mu - \hat{\mu}}{\sigma_\mu} \right)^2 \quad . \tag{8.6}$$

This allows to directly calculate p_0 and the significance. The results provided by `TRExFitter` are:

$$p_{0,\mathrm{exp}} = 0.322 \qquad , \qquad Z_{0,\mathrm{exp}} = 0.461 \tag{8.7}$$

$$p_{0,\mathrm{obs}} = 0.588 \qquad , \qquad Z_{0,\mathrm{obs}} = -0.222 \ . \tag{8.8}$$

Hence, by using only the ATLAS 2017 data, a 0.461σ deviation in the $ZZ \to 4\tau$ cross section from the Standard Model prediction is expected. Due to the negative fit result for $\mu(ZZ \to 4\tau)$, the observed deviation is negative as well, namely -0.222σ. Although these values are far away from a 3σ excess or even a discovery at 5σ, they show the potential of an approach where a dedicated identification method for multi-tau events is combined with the power of boosted decision trees in order to separate the few expected signal events from an extremely large background. The use of the full ATLAS run 2 data with an integrated luminosity of 139 fb^{-1} probably would allow to further improve this result.

[3] This theorem even holds for a data sample with a size of $\approx O(10)$ events [72], so it is also applicable in this analysis where the event yields in the signal enriched bins are quite low.

CHAPTER 9

Conclusion

In this work the potential of different diboson processes decaying into 4τ final states using ATLAS 2017 data with an integrated luminosity of $46.9\,\mathrm{fb}^{-1}$ was probed.
For that purpose, at first the power of a dedicated identification method for multi-tau events developed in [8] was demonstrated in chapter 5 using a $hh \to 4\tau$ signal sample again. In contrast to the usual ATLAS approach, where every single τ-lepton needs to fulfill a fixed working point corresponding to a certain τ^{jetBDT} score threshold for an event to be accepted, the method of multiplied τ-lepton probabilities assigns a combined probability score $\log\left(\prod \tau^{\mathrm{prob}}_{\mathrm{i=1\text{-}3,4}}\right)$ to each event. This score then serves as the selection criterion. Figure 5.7 shows a significant improvement in the achievable background rejection, especially for fake tau enriched background processes like $t\bar{t}$ production. These observations match the results from the studies already performed in [8].
In chapter 6, the expected background in the 4τ channel was studied. A data driven estimation technique based on combinatorial considerations of the reconstructed four τ-lepton charges for fake tau events stemming from QCD multijet production [8] (cf. fig. 6.3) has been adapted to further account for fake tau events originating from heavy particle decays. They are now estimated from data as well instead of being simulated using Monte Carlo event generators. This advanced fake estimation method provides a good data-to-model agreement (cf. fig. 6.5) and makes the technique more stable in the phase space region of high $\log\left(\prod \tau^{\mathrm{prob}}_{\mathrm{i=1\text{-}3,4}}\right)$ scores.
After these preparatory studies, one of the main goals of this work was to verify the introduced methods on a Standard Model process. Therefore, a $ZZ \to 4\tau$ sample was extracted from a Sherpa $\ell\ell\ell\ell$ sample which then served as signal for the remaining analysis. First, the outcome of a cut-based selection of $ZZ \to 4\tau$ events using the tau probability product $\log\left(\prod \tau^{\mathrm{prob}}_{\mathrm{i=1\text{-}3,4}}\right)$ was evaluated in section 7.1. Although this approach already allows to reject a large fraction of the expected background, the gain provided by $\log\left(\prod \tau^{\mathrm{prob}}_{\mathrm{i=1\text{-}3,4}}\right)$ on its own is limited due to the low absolute number of signal events (cf. fig. 7.2). Nevertheless, a significance of up to $S/\sqrt{S+B} = 0.635$ can be reached by combining the tau probability product with an additional veto against b-jets (cf. tab. 7.1).
In order to improve this result, a multivariate event selection approach was taken in section 7.2. Two forests of boosted decision trees were trained to discriminate the $ZZ \to 4\tau$ signal from QCD multijet and fake tau events originating from heavy particle decays, respectively. A careful optimization of the training variables and the hyperparameter setup was performed for both of them. The final training parameters are summarized in the tables 7.2 and 7.3, while the obtained evaluation of the

input variables ranked by their importance for discrimination is shown in table 7.4 and 7.5. During the optimization process, a balancing between maximizing the separation performance and keeping overtraining effects at an acceptable level had to be found. With the finally chosen setup, AUC values of 0.803 for the Monte Carlo BDT and 0.909 for the QCD multijet BDT are obtained when being evaluated on the total background (cf. tab. 7.9, but also fig. 7.8 and 7.10). The classifier output distributions in figure 7.7 and 7.9 especially indicate a good discrimination performance against the background events the respective classifier has been trained on.
In a next step, these multivariate classifiers were then combined with the power of the tau probability product to enhance the selection of $ZZ \to 4\tau$ events. By binning in both of them and the $N_{\text{b-jets}}$ observable, overall six phase space regions were constructed (cf. tab. 7.10). Four of them contain the majority of the expected background, while the other two are enriched with signal events at high $\log\left(\prod \tau^{\text{prob}}_{\text{i=1-3,4}}\right)$ values (cf. fig. 7.13, 7.15 and 7.16).
In chapter 8, these regions were finally used as input for a profile likelihood fit carried out with `TRExFitter` [70], in order to measure the $ZZ \to 4\tau$ signal strength compared to the Standard Model prediction. The treatment of systematic uncertainties related to tau identification was discussed in section 8.2.1 before further nuisance parameters on the estimated backgrounds were motivated in section 8.2.2. Then, first an Asimov dataset was used to test the technical fit setup. Afterwards, the fit was performed on the ATLAS 2017 data. For the parameter of interest, $\mu(ZZ \to 4\tau)$, a value of $-0.49^{+2.25}_{-1.59}$ is obtained. For the uncertainty on the Monte Carlo background, a nuisance parameter pull is observed, but can be explained from the event distributions in the different fit regions. The effect of further tau ID related systematic uncertainties is found to be very small (cf. tab. 8.2). Lastly, an expected limit of 5.488 times the Standard Model prediction can be set on the $ZZ \to 4\tau$ signal strength and a significance of 0.461σ is calculated.

One of the main limitations during this work was the limited number of signal events for the BDT training process. As visible in the plots evaluating different hyperparameter setups in appendix D, the separation between signal and background events for higher tree depths and smaller minimum node sizes is much better on the training sample than on the test sample. This leads to the conclusion that the performance could be further enhanced if a sufficiently large training sample would be available. An improvement in the BDT training should have a direct impact on the results obtained in the profile likelihood fit. In a next step, the analysis could be repeated using the full ATLAS run 2 data with an integrated luminosity of 139 fb^{-1} which would reduce the limitations currently caused by low statistics. Furthermore, additional systematic uncertainties need to be included. On the long time scale, the High-Luminosity LHC will provide a significantly larger amount of data and therefore of $ZZ \to 4\tau$ events, too.
Finally, as already motivated at the begin of this thesis, the exotic final state of four τ-leptons also has some potential to probe theories beyond the Standard Model. A further investigation of this area, focussing e.g. on BSM resonances like heavy Higgs bosons as predicted by two-Higgs doublet models [78], which then decay into two on-shell Z bosons, might thus be interesting.

Bibliography

[1] Alexander von Humboldt, Kosmos. Entwurf einer physischen Weltbeschreibung, 5 Bde., 1845-62. Taken from https://www.aphorismen.de/zitat/88827; (retrieved 27.09.2021) (cit. on p. 1).

[2] I. Falconer, *J J Thomson and the discovery of the electron*, Physics Education **32** (1997) 226, URL: https://doi.org/10.1088/0031-9120/32/4/015 (cit. on p. 1).

[3] A. Einstein, *On the Electrodynamics of Moving Bodies*, online available at http://hermes.ffn.ub.es/luisnavarro/nuevo_maletin/Einstein_1905_relativity.pdf; (retrieved 27.09.2021), 1905 (cit. on p. 1).

[4] J. D. Norton, *Einstein's Special Theory of Relativity and the Problems in the Electrodynamics of Moving Bodies that led him to it.* Online available at https://sites.pitt.edu/~jdnorton/papers/companion.pdf; (retrieved 27.09.2021), 2004 (cit. on p. 1).

[5] N. Bohr, *On the Constitution of Atoms and Molecules*, Philosophical Magazine **26** (1913) 1, online available at http://hermes.ffn.ub.es/luisnavarro/nuevo_maletin/Bohr_1913.pdf; (retrieved 27.09.2021) (cit. on p. 1).

[6] E. Schrödinger, *An Undulatory Theory of the Mechanics of Atoms and Molecules*, The Physical Review **28** (1926), online available at https://people.isy.liu.se/jalar/kurser/QF/references/Schrodinger1926c.pdf; (retrieved 27.09.2021) (cit. on p. 1).

[7] G. Aad et al., *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Physics Letters B **716** (2012) 1, ISSN: 0370-2693, URL: http://dx.doi.org/10.1016/j.physletb.2012.08.020 (cit. on pp. 1, 3).

[8] H. Junkerkalefeld, *Study Towards Analysing the Higgs Self-Coupling in the $hh \to 4\tau$ Channel at the ATLAS Detector*, online available at https://www.lhc-ilc.physik.uni-bonn.de/ergebnisse/dateien/t00000115.pdf?c=t&id=115; (retrieved 26.08.2021), M.Sc. Thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, 2019 (cit. on pp. 1, 2, 33, 34, 37–39, 42–47, 52, 74, 87, 97).

[9] G. Altarelli and M. W. Grünewald, *Precision electroweak tests of the Standard Model*, Physics Reports **403-404** (2004) 189, ISSN: 0370-1573, URL: http://dx.doi.org/10.1016/j.physrep.2004.08.013 (cit. on p. 3).

[10] M. Thomson, *Modern Particle Physics*, 6th ed., Cambridge University Press, 2013, ISBN: 978-1-107-03426-6 (cit. on pp. 3, 5–14).

[11] B. Povh, K. Rith et al., *Teilchen und Kerne*, 9th ed., Springer-Verlag Berlin Heidelberg, 2014, ISBN: 978-3-642-37822-5 (cit. on pp. 3–5, 15).

[12] Online available at https://i.insider.com/59515711d084cc817f8b4ebc?width=1000&format=jpeg&auto=webp; (retrieved 16.08.2021) (cit. on p. 4).

[13] M. Srednicki, *Quantum Field Theory*, online available at https://web.physics.ucsb.edu/~mark/ms-qft-DRAFT.pdf; (retrieved 18.08.2021) (cit. on p. 5).

[14] *Created with the TikZ package in* LaTeX (cit. on pp. 5, 11, 13).

[15] S. Weinberg, *Conceptual Foundations of the Unified Theory of Weak and Electromagnetic Interactions*, online available at https://www.nobelprize.org/uploads/2018/06/weinberg-lecture.pdf; (retrieved 18.08.2021), 1979 (cit. on p. 6).

[16] G. Arnison et al., *Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c**2 at the CERN SPS Collider*, Phys. Lett. B **126** (1983) 398, URL: https://inspirehep.net/literature/190338 (cit. on p. 8).

[17] L. Di Lella and C. Rubbia, *The Discovery of the W and Z Particles*, online available at https://cds.cern.ch/record/2103277/files/9789814644150_0006.pdf?subformat=pdfa&version=1; (retrieved 19.08.2021) (cit. on p. 8).

[18] M. Tabanashi et al. (Particle Data Group), Phys. Rev. D **98** (2018) 030001, URL: https://pdg.lbl.gov/2018/listings/rpp2018-list-z-boson.pdf (cit. on pp. 8, 9).

[19] S. Richter, *Measurements of ZZ production with the ATLAS detector and simulation of loop-induced processes with the HERWIG event generator*, (2018), online available at https://arxiv.org/pdf/1806.02697.pdf; (retrieved 20.08.2021), arXiv: 1806.02697 [hep-ex] (cit. on pp. 9, 10).

[20] J. Wang, *Diboson production at LHC and Tevatron*, International Journal of Modern Physics: Conference Series **31** (2014) 1460279, ISSN: 2010-1945, URL: http://dx.doi.org/10.1142/S2010194514602798 (cit. on p. 10).

[21] Online available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/SM/ATLAS_a_SMSummary_TotalXsect/ATLAS_a_SMSummary_TotalXsect.png; (retrieved 20.08.2021) (cit. on p. 10).

[22] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (16 1964) 508, URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.508 (cit. on p. 11).

[23] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (9 1964) 321, URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.321 (cit. on p. 11).

[24] Online available at https://www.researchgate.net/profile/Avelino-Vicente/publication/50991472/figure/fig1/AS:669459207426055@1536622921553/The-well-known-SM-Higgs-potential-for-two-parameter-configurations-If-l-0-and-2.jpg; (retrieved 23.08.2021) (cit. on p. 12).

[25] Online available at https://physicstoday.scitation.org/na101/home/literatum/publisher/aip/journals/content/pto/2013/pto.2013.66.issue-12/pt.3.2212/production/images/medium/pt.3.2212.figures.f2.gif; (retrieved 23.08.2021) (cit. on p. 12).

[26] M. Tabanashi et al. (Particle Data Group), Phys. Rev. D **98** (2018) 030001, URL: https://pdg.lbl.gov/2019/tables/rpp2019-sum-gauge-higgs-bosons.pdf (cit. on p. 13).

[27] K.A. Olive et al. (Particle Data Group), Chin. Phys. C **38** (2014) 090001, URL: https://pdg.lbl.gov/2014/listings/rpp2014-list-tau.pdf (cit. on pp. 13, 33).

[28] Online available at http://opendata.atlas.cern/books/current/get-started/_book/the-higgs-boson.html; (retrieved 23.08.2021) (cit. on p. 14).

[29] T. Rashid, *Development of pixel detector for ATLAS Inner Tracker(ITK) upgrade at HL-LHC and Searching for the Standard Model Higgs boson decay into b-quark pair with ATLAS experiment*, Online available at https://www.researchgate.net/publication/333942621_Development_of_pixel_detector_for_ATLAS_Inner_TrackerITK_upgrade_at_HL-LHC_and_Searching_for_the_Standard_Model_Higgs_boson_decay_into_b-quark_pair_with_ATLAS_experiment; (retrieved 28.10.2021), PhD thesis, 2019 (cit. on p. 14).

[30] Education, Communications and Outreach Group CERN, *LHC - the guide*, online available at https://home.cern/sites/home.web.cern.ch/files/2018-07/CERN-Brochure-2017-002-Eng_0.pdf; (retrieved 03.08.2021) (cit. on pp. 15, 16).

[31] CERN, *High-Luminosity LHC*, online available at https://home.cern/resources/faqs/high-luminosity-lhc; (retrieved 03.08.2021) (cit. on p. 15).

[32] CERN, *LHC Season 2 - facts & figures*, online available at https://home.cern/sites/home.web.cern.ch/files/2018-07/factsandfigures-en_0.pdf; (retrieved 03.08.2021) (cit. on p. 16).

[33] Online available at https://cds.cern.ch/record/2684277/files/CCC-v2019-final-white.png?subformat=icon-640; (retrieved 04.08.2021) (cit. on p. 16).

[34] The ATLAS Collaboration et al., *The ATLAS Experiment at the CERN Large Hadron Collider*, Journal of Instrumentation **3** (2008) S08003, URL: https://doi.org/10.1088/1748-0221/3/08/s08003 (cit. on pp. 16–21).

[35] The ATLAS Collaboration, *Detector & Technology*, online available at https://atlas.cern/discover/detector; (retrieved 04.08.2021) (cit. on p. 17).

[36] Online available at https://www.researchgate.net/publication/279019406/figure/fig9/AS:668867969961989@1536481959229/Schematic-of-the-ATLAS-detector.ppm; (retrieved 04.08.2021) (cit. on p. 17).

[37] I. Vichou, on behalf of the ATLAS Collaboration, *Performance of Missing Transverse Momentum Reconstruction in ATLAS with Proton-Proton Collisions at $\sqrt{s}$ = 7 TeV*, Journal of Physics: Conference Series **404** (2012) 012011, URL: https://doi.org/10.1088/1742-6596/404/1/012011 (cit. on p. 18).

[38] Online available at https://cds.cern.ch/record/1699952/files/Figures_T_Coordinate.png; (retrieved 04.08.2021) (cit. on p. 18).

[39] Online available at https://mediastream.cern.ch/MediaArchive/Photo/Public/2008/0803014/0803014_01/0803014_01-A5-at-72-dpi.jpg; (retrieved 04.08.2021) (cit. on p. 19).

[40] Online available at https://mediaarchive.cern.ch/MediaArchive/Photo/Public/2008/0803015/0803015_01/0803015_01-A4-at-144-dpi.jpg; (retrieved 05.08.2021) (cit. on p. 20).

[41] C. Berger, *Elementarteilchenphysik*, 3rd ed., Springer-Verlag Berlin Heidelberg, 2014, ISBN: 978-3-642-41752-8 (cit. on p. 21).

[42] S. Xella, on behalf of the ATLAS collaboration, *Physics object reconstruction in the ATLAS experiment*, PoS **CHARGED2012** (2012) 011, ed. by R. Enberg and A. Ferrari (cit. on pp. 21, 23).

[43] M. Cacciari, G. P. Salam and G. Soyez, *The anti-kt jet clustering algorithm*, Journal of High Energy Physics **2008** (2008) 063, URL: https://doi.org/10.1088/1126-6708/2008/04/063 (cit. on pp. 21, 33).

[44] The ATLAS Collaboration, *Jet energy scale and its systematic uncertainty in proton-proton collisions at $\sqrt{s}$ = 7 TeV in ATLAS 2010 data*, tech. rep., online available at https://cds.cern.ch/record/1337782; (retrieved 05.08.2021): CERN, 2011 (cit. on p. 21).

[45] Online available at https://cds.cern.ch/record/1519066/files/Figure_003-a.png; (retrieved 05.08.2021) (cit. on p. 22).

[46] A. Buckley et al., *General-purpose event generators for LHC physics*, Physics Reports **504** (2018), arXiv: 1101.2599v1 [hep-ph] (cit. on pp. 23, 24).

[47] S. Ferreres-Solé and T. Sjöstrand, *The Space-Time Structure of Hadronization in the Lund Model*, (2018), online available at https://arxiv.org/pdf/1808.04619.pdf; (retrieved 06.08.2021), arXiv: 1808.04619v2 [hep-ph] (cit. on p. 23).

[48] T. Sjöstrand, S. Mrenna and P. Skands, *A brief introduction to PYTHIA 8.1*, Computer Physics Communications **178** (2008) 852, ISSN: 0010-4655, URL: http://dx.doi.org/10.1016/j.cpc.2008.01.036 (cit. on p. 24).

[49] The ATLAS Collaboration et al., *The ATLAS Simulation Infrastructure*, The European Physical Journal C (2010), URL: https://doi.org/10.1140/epjc/s10052-010-1429-9 (cit. on p. 24).

[50] T. Gleisberg et al., *SHERPA 1. , a proof-of-concept version*, Journal of High Energy Physics **2004** (2004) 056, ISSN: 1029-8479, URL: http://dx.doi.org/10.1088/1126-6708/2004/02/056 (cit. on pp. 24, 43).

[51] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, Journal of High Energy Physics **2007** (2007) 070, ISSN: 1029-8479, URL: http://dx.doi.org/10.1088/1126-6708/2007/11/070 (cit. on p. 24).

[52] S. Frixione, P. Nason and B. R. Webber, *Matching NLO QCD and parton showers in heavy flavour production*, Journal of High Energy Physics **2003** (2003) 007, ISSN: 1029-8479, URL: http://dx.doi.org/10.1088/1126-6708/2003/08/007 (cit. on pp. 24, 43).

[53] S. Agostinelli, J. Allison, K. Amako et al., *Geant4—a simulation toolkit*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **506** (2003) 250, ISSN: 0168-9002, URL: https://www.sciencedirect.com/science/article/pii/S0168900203013688 (cit. on p. 24).

[54] S. Höche, *Introduction to parton-shower event generators*, (2015), online available at https://arxiv.org/pdf/1411.4085.pdf; (retrieved 05.08.2021), arXiv: 1411.4085v2 [hep-ph] (cit. on p. 25).

[55] K. Albertsson, S. Gleyzer, A. Hoecker et al., *TMVA 4 Users Guide*, online available at https://root.cern.ch/download/doc/tmva/TMVAUsersGuide.pdf; (retrieved 10.08.2021) (cit. on pp. 27–32).

[56] R. Brun and F. Rademakers, *ROOT - An Object Oriented Data Analysis Framework*, Nucl. Inst. & Meth. in Phys. Res. A **389** (1997), see also https://root.cern/ (cit. on pp. 27, 77).

[57] V. Nasteski, *An overview of the supervised machine learning methods*, HORIZONS.B **4** (2017) 51 (cit. on p. 27).

[58] Online available at https://www.researchgate.net/publication/286016178/figure/fig7/AS:881192740081666@1587104129288/color-online-Distributions-of-the-BDT-classifier-for-both-training-and-test-samples-of.png; (retrieved 11.08.2021) (cit. on p. 29).

[59] Y. Coadou, *Boosted Decision Trees and Applications*, EPJ Web of Conferences **55** (2013) 02004 (cit. on p. 29).

[60] B. Mehlig, *Machine learning with neural networks*, online available at https://arxiv.org/pdf/1901.05639.pdf; (retrieved 11.08.2021), 2021, arXiv: 1901.05639 [cs.LG] (cit. on p. 29).

[61] J. Therhaag, *Introduction to neural networks in high energy physics*, EPJ Web of Conferences **55** (2013) 02003 (cit. on p. 29).

[62] Brownlee, Jason, *Loss and Loss Functions for Training Deep Learning Neural Networks*, online available at https://machinelearningmastery.com/loss-and-loss-functions-for-training-deep-learning-neural-networks/; (retrieved 13.08.2021) (cit. on p. 31).

[63] The ATLAS Collaboration, *Identification and energy calibration of hadronically decaying tau leptons with the ATLAS experiment in pp collisions at* $\sqrt{s} = 8$ TeV, The European Physical Journal C **75** (2015), URL: https://link.springer.com/content/pdf/10.1140/epjc/s10052-015-3500-z.pdf (cit. on p. 33).

[64] The ATLAS Collaboration, *Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC*, ATLAS-PHYS-PUB-2015-045, online available at https://cds.cern.ch/record/2064383/files/ATL-PHYS-PUB-2015-045.pdf; (retrieved 27.08.2021), 2015 (cit. on pp. 33–36).

[65] The ATLAS Collaboration, *Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment*, ATLAS-PHYS-PUB-2019-033, online available at https://cds.cern.ch/record/2688062/files/ATL-PHYS-PUB-2019-033.pdf; (retrieved 30.08.2021), 2019 (cit. on pp. 34, 37).

[66] J. Bellm et al., *Herwig 7.0 / Herwig++ 3.0 release note*, The European Physical Journal C **76** (2016), ISSN: 1434-6052, URL: http://dx.doi.org/10.1140/epjc/s10052-016-4018-8 (cit. on p. 43).

[67] A. de Maria, E. Drechsler and T. Zorbas, *xTauFramework*, online available at https://gitlab.cern.ch/ATauLeptonAnalysiS/xTauFramework; (retrieved 15.09.2021) (cit. on p. 44).

[68] *MET Significance*, online available at https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideMETSignificance; (retrieved 25.10.2021) (cit. on p. 55).

[69] M. Hohenwarter et al., *GeoGebra 4.4*, online available at https://www.geogebra.org/; (retrieved 26.09.2021), 2013 (cit. on p. 66).

[70] *TRExFitter*, online available at https://gitlab.cern.ch/TRExStats/TRExFitter (ATLAS internal); (retrieved 12.10.2021) (cit. on pp. 73, 88).

[71] K. Schubert, *Application of the Profile-Likelihood Method to the Combination of Cross Section Measurements*, online available at http://iktp.tu-dresden.de/IKTP/pub/12/bachelor_KS.pdf; (retrieved 07.10.2021), B.Sc. Thesis: Technische Universität Dresden, 2012 (cit. on p. 73).

[72] M. Pinamonti, *Statistical Methods at ATLAS and CMS*, Higgs Toppings Workshop, Benasque, 28th May - 1st June 2018, online available at https://indico.cern.ch/event/727396/contributions/3021899/attachments/1657532/2654085/Statistical_methods_at_ATLAS_and_CMS_2.pdf; (retrieved 07.10.2021) (cit. on pp. 73, 74, 85).

[73] The ATLAS Collaboration, *Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at* $\sqrt{s} = 13$ TeV, ATLAS-CONF-2017-029, online available at http://cdsweb.cern.ch/record/2261772/files/ATLAS-CONF-2017-029.pdf; (retrieved 09.10.2021), 2017 (cit. on p. 75).

[74] Online available at https://twiki.cern.ch/twiki/pub/AtlasProtected/TauRecommendationsR21/ValidationPlots-2019-summer.pdf; (retrieved 09.10.2021) (cit. on p. 75).

[75] L. Gerlach and S. Lai, *Update on* τ *ID Scale Factors for BDT & RNN*, online available at https://indico.cern.ch/event/830580/contributions/3595031/attachments/1923123/3182036/tauid-19-10-08.pdf; (retrieved 09.10.2021), 2019 (cit. on p. 75).

[76] *The* CL_s *method; information for conference speakers*, online available at https://www.pp.rhul.ac.uk/~cowan/stat/cls/CLsInfo.pdf; (retrieved 11.10.2021), 2011 (cit. on pp. 78, 84).

[77] A. L. Read, *Presentation of search results: the* CL_s *technique*, Journal of Physics G: Nuclear and Particle Physics **28** (2002) 2693, URL: https://doi.org/10.1088/0954-3899/28/10/313 (cit. on pp. 78, 84).

[78] G. Branco et al., *Theory and phenomenology of two-Higgs-doublet models*, Physics Reports **516** (2012) 1, ISSN: 0370-1573, URL: http://dx.doi.org/10.1016/j.physrep.2012.02.002 (cit. on p. 88).

APPENDIX A

Detailed List of Used Monte Carlo Samples

This chapter lists the different Monte Carlo samples used in the analysis performed in this thesis. The nomenclature is copied from [8] since exactly the same samples (on n-tuple level) are used.

Dataset ID	Name	Generator	σ_{eff} [pb]
342284	WH125 inc	Pythia8EvtGen	1.38
342285	ZH125 inc	Pythia8EvtGen	0.87
344235	VBFH125 ZZ4lep notau	PowhegPythia8EvtGen	3.49
345060	ggH125 ZZ4l	PowhegPythia8EvtGen	0.013
345066	ggZH125 ZZ4lepZinc	PowhegPythia8EvtGen	0.00024
345123	ggH125 tautau	PowhegPythia8EvtGen	1.37
345211	WmH125J Winc tautau	PowhegPythia8EvtGen	0.033
345212	WpH125J Winc tautau	PowhegPythia8EvtGen	0.052
345217	ZH125J Zinc tautau	PowhegPythia8EvtGen	0.055
345875	ttH125 dilep	PowhegPythia8EvtGen	0.062
345948	VBFH125 WWlvlv	PowhegPythia8EvtGen	0.086
346193	VBFH125 tautauh	PowhegPythia8EvtGen	0.11
345911	hh 4tau	aMcAtNloHerwig7EvtGen	0.00012

Table A.1: MC samples grouped into the Higgs background process. In the second block, the $hh \rightarrow 4\tau$ signal sample used in chapter 5 and 6 is given.

Dataset ID	Name	Generator	σ_{eff} [pb]
410470	ttbar nonallhad	PhPy8EG	396.88
410471	ttbar allhad	PhPy8EG	332.98
410472	ttbar dil	PhPy8EG	76.96

Table A.2: MC samples grouped into the $t\bar{t}$ background process.

Dataset ID	Name	Generator	σ_{eff} [pb]
410644	singletop schan lept top	PowhegPythia8EvtGen	2.027
410645	singletop schan lept antitop	PowhegPythia8EvtGen	1.267
410646	Wt DR inclusive top	PowhegPythia8EvtGen	37.94
410647	Wt DR inclusive antitop	PowhegPythia8EvtGen	37.91
410654	Wt DS inclusive top	PowhegPythia8EvtGen	36.92
410655	Wt DS inclusive antitop	PowhegPythia8EvtGen	37.66

Table A.3: MC samples grouped into the Top background process.

Dataset ID	Name	Generator	σ_{eff} [pb]
364128	Ztautau p_{T} range: 0 70 CVetoBVeto	Sherpa221	1587.20
364129	Ztautau p_{T} range: 0 70 CFilterBVeto	Sherpa221	218.31
364130	Ztautau p_{T} range: 0 70 BFilter	Sherpa221	124.55
364131	Ztautau p_{T} range: 70 140 CVetoBVeto	Sherpa221	74.13
364132	Ztautau p_{T} range: 70 140 CFilterBVeto	Sherpa221	19.71
364133	Ztautau p_{T} range: 70 140 BFilter	Sherpa221	11.99
364134	Ztautau p_{T} range: 140 280 CVetoBVeto	Sherpa221	24.19
364135	Ztautau p_{T} range: 140 280 CFilterBVeto	Sherpa221	9.10
364136	Ztautau p_{T} range: 140 280 BFilter	Sherpa221	5.34
364137	Ztautau p_{T} range: 280 500 CVetoBVeto	Sherpa221	4.67
364138	Ztautau p_{T} range: 280 500 CFilterBVeto	Sherpa221	2.22
364139	Ztautau p_{T} range: 280 500 BFilter	Sherpa221	1.47
364140	Ztautau p_{T} range: 500 1000	Sherpa221	1.76
364141	Ztautau p_{T} range: 1000 E CMS	Sherpa221	0.14

Table A.4: MC samples grouped into the $Z \to \tau\tau$ background process. They are sliced into different p_{T} ranges and several flavor filters are applied.

Dataset ID	Name	Generator	σ_{eff} [pb]
308092	Zee2jets TChannel	Sherpa 221	0.63
308093	Zmm2jets TChannel	Sherpa 221	0.63
308094	Ztautau2jets TChannel	Sherpa 221	0.63
308095	Znunu2jets TChannel	Sherpa 221	2.94
308096	Wenu2jets TChannel	Sherpa 221	6.81
308097	Wmunu2jets TChannel	Sherpa 221	6.80
308098	Wtaunu2jets TChannel	Sherpa 221	6.80
364100	Zmumu p_{T} range: 0 70 CVetoBVeto	Sherpa 221	1589.63
364101	Zmumu p_{T} range: 0 70 CFilterBVeto	Sherpa 221	218.15
364102	Zmumu p_{T} range: 0 70 BFilter	Sherpa 221	124.01
364103	Zmumu p_{T} range: 70 140 CVetoBVeto	Sherpa 221	73.15
364104	Zmumu p_{T} range: 70 140 CFilterBVeto	Sherpa 221	19.84
364105	Zmumu p_{T} range: 70 140 BFilter	Sherpa 221	12.08
364106	Zmumu p_{T} range: 140 280 CVetoBVeto	Sherpa 221	23.68
364107	Zmumu p_{T} range: 140 280 CFilterBVeto	Sherpa 221	9.04
364108	Zmumu p_{T} range: 140 280 BFilter	Sherpa 221	5.75
364109	Zmumu p_{T} range: 280 500 CVetoBVeto	Sherpa 221	4.65
364110	Zmumu p_{T} range: 280 500 CFilterBVeto	Sherpa 221	2.21
364111	Zmumu p_{T} range: 280 500 BFilter	Sherpa 221	1.45
364112	Zmumu p_{T} range: 500 1000	Sherpa 221	1.74
364113	Zmumu p_{T} range: 100 E CMS	Sherpa 221	0.14
364114	Zee p_{T} range: 0 70 CVetoBVeto	Sherpa 221	1586.66
364115	Zee p_{T} range: 0 70 CFilterBVeto	Sherpa 221	218.16
364116	Zee p_{T} range: 0 70 BFilter	Sherpa 221	123.30
364117	Zee p_{T} range: 70 140 CVetoBVeto	Sherpa 221	74.39
364118	Zee p_{T} range: 70 140 CFilterBVeto	Sherpa 221	19.83
364119	Zee p_{T} range: 70 140 BFilter	Sherpa 221	12.31
364120	Zee p_{T} range: 140 280 CVetoBVeto	Sherpa 221	24.41
364121	Zee p_{T} range: 140 280 CFilterBVeto	Sherpa 221	9.14
364122	Zee p_{T} range: 140 280 BFilter	Sherpa 221	5.93
364123	Zee p_{T} range: 280 500 CVetoBVeto	Sherpa 221	4.75
364124	Zee p_{T} range: 280 500 CFilterBVeto	Sherpa 221	2.22
364125	Zee p_{T} range: 280 500 BFilter	Sherpa 221	1.46
364126	Zee p_{T} range: 500 1000	Sherpa 221	1.76
364127	Zee p_{T} range: 1000 E CMS	Sherpa 221	0.14

Table A.5: MC samples grouped into the V + Jets background process. The first block summarizes different t-channel processes, while the second and third one contain $Z \to ee$ and $Z \to \mu\mu$ s-channel events. These are sliced into different p_{T} ranges and several flavor filters are applied.

Dataset ID	Name	Generator	σ_{eff} [pb]
363355	ZqqZvv	Sherpa 221	4.35
363356	ZqqZll	Sherpa 221	2.17
363357	WqqZvv	Sherpa 221	6.80
363358	WqqZll	Sherpa 221	3.44
363359	WpqqWmlv	Sherpa 221	24.72
363360	WplvWmqq	Sherpa 221	24.73
363489	WlvZqq	Sherpa 221	11.41
364250	llll	Sherpa 222	1.25
364253	lllv	Sherpa 222	4.58
364254	llvv	Sherpa 222	12.50
364255	lvvv	Sherpa 222	3.23
364283	lllljj	Sherpa 222	0.01
364284	lllvjj	Sherpa 222	0.05
364285	llvvjj	Sherpa 222	0.12
364288	llll lowMllPtComplement	Sherpa 222	1.43
364289	lllv lowMllPtComplement	Sherpa 222	2.91
364290	llvv lowMllPtComplement	Sherpa 222	0.17
407311	6l0v	Sherpa 221	0.00010
407312	5l1v	Sherpa 221	0.00057
407313	4l2v	Sherpa 221	0.0044
407314	3l3v	Sherpa 221	0.016
407315	2l4v	Sherpa 221	0.0058

Table A.6: MC samples grouped into the Multiboson background process. The upper two blocks contain diboson processes, while the bottom one consists of triboson processes. The $ZZ \to 4\tau$ process used in chapter 7 and 8 is part of the llll sample which contains all SM processes with exactly four leptons in the final state. The $hh \to 4\tau$ process is not included due to its limited statistics.

APPENDIX B

Input Variable Distributions for the BDT Training

The transformed distributions of all input variables used for the BDT training, which are not shown in section 7.2.1, can be found below.

Monte Carlo BDT

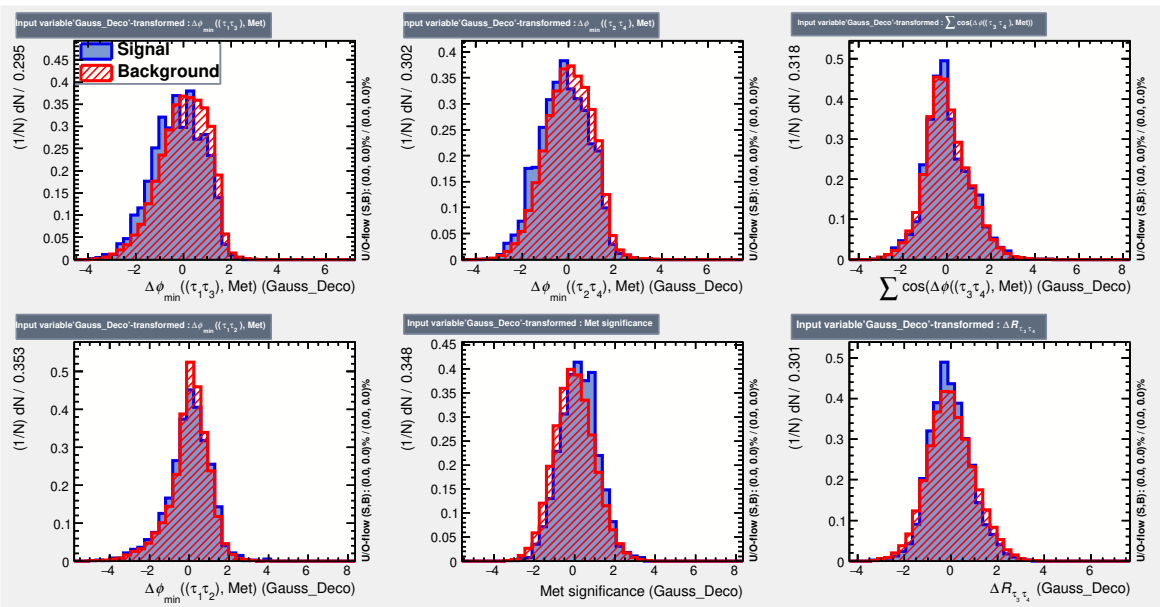


Figure B.1: Further input variables for the Monte Carlo BDT training (1/7).

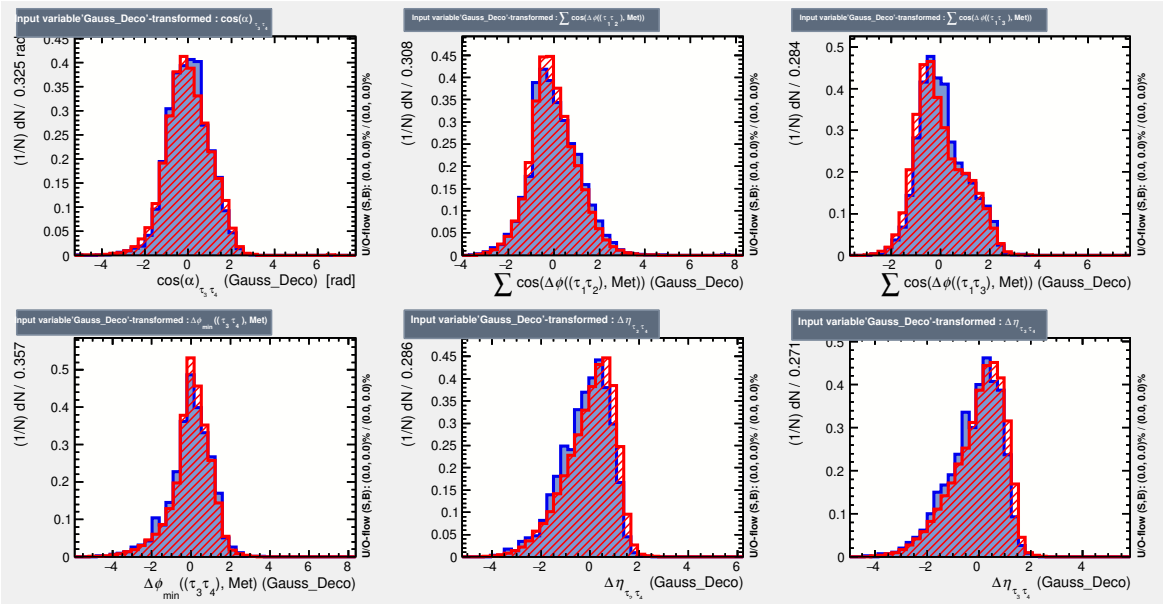


Figure B.2: Further input variables for the Monte Carlo BDT training (2/7).

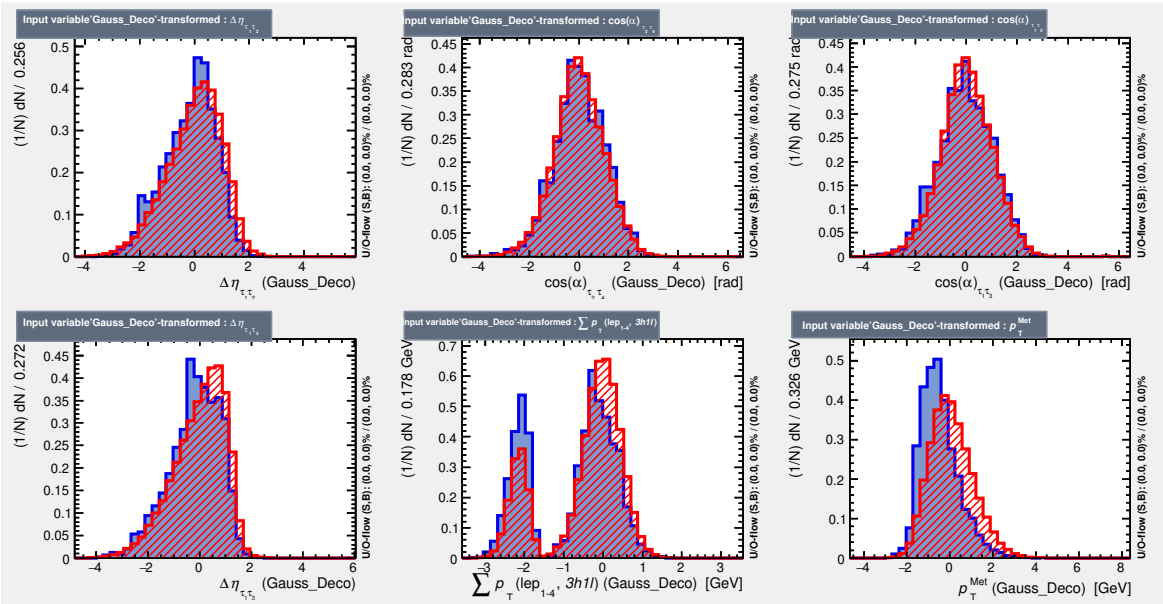


Figure B.3: Further input variables for the Monte Carlo BDT training (3/7).

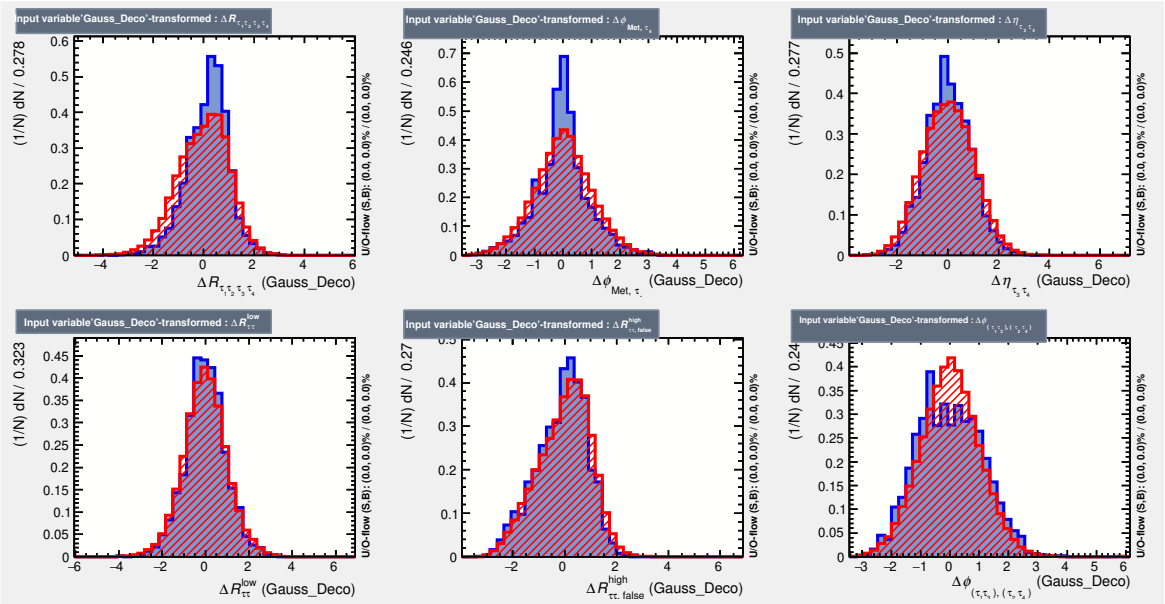


Figure B.4: Further input variables for the Monte Carlo BDT training (4/7).

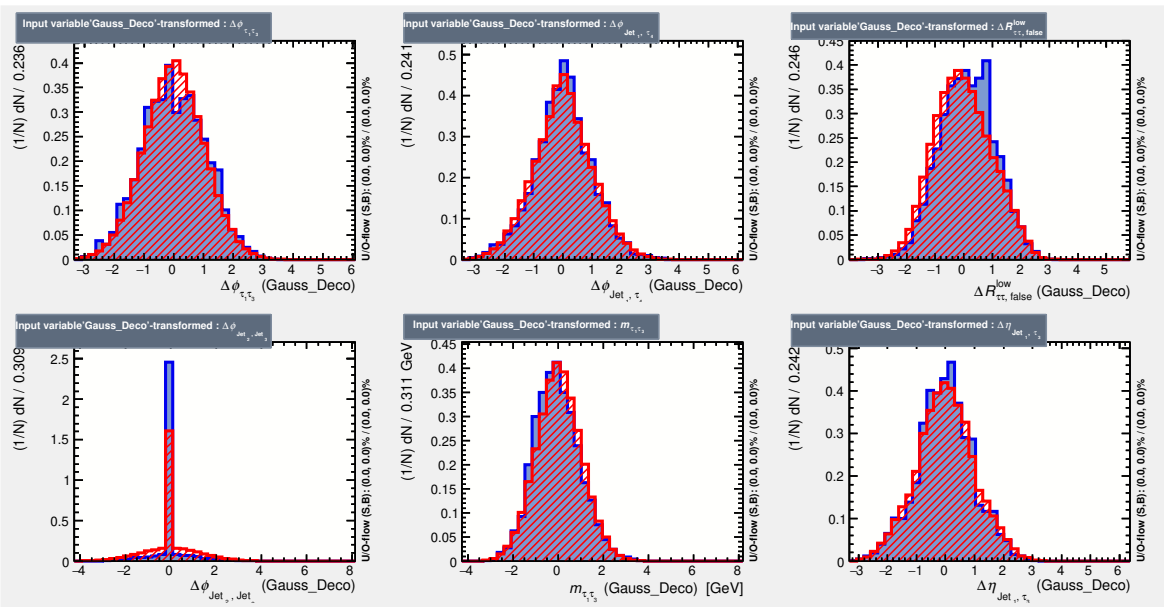


Figure B.5: Further input variables for the Monte Carlo BDT training (5/7).

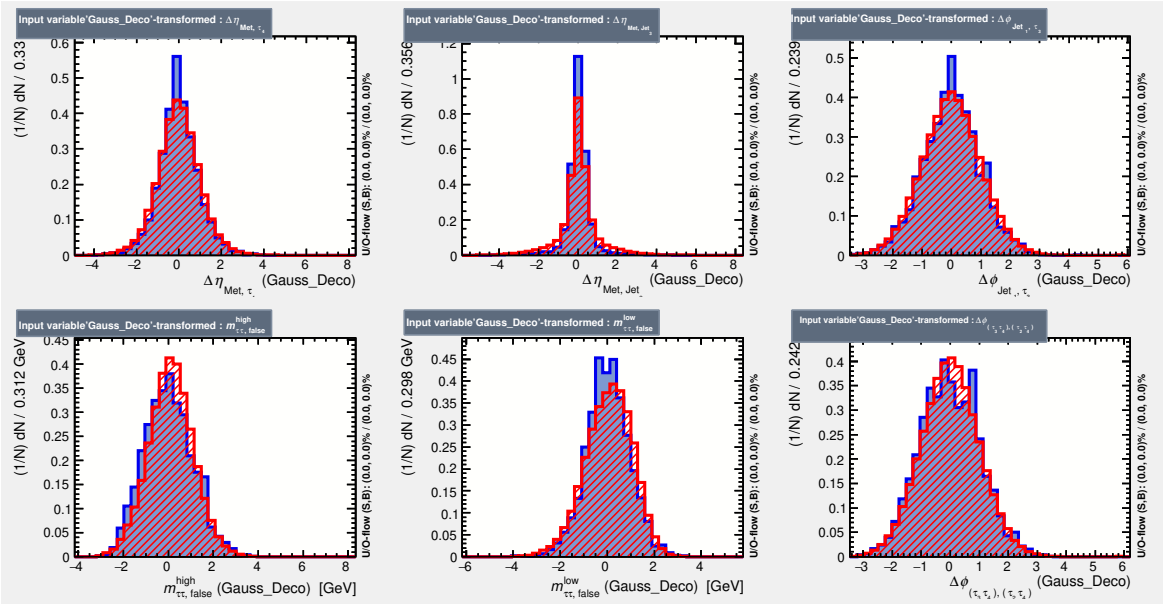


Figure B.6: Further input variables for the Monte Carlo BDT training (6/7).

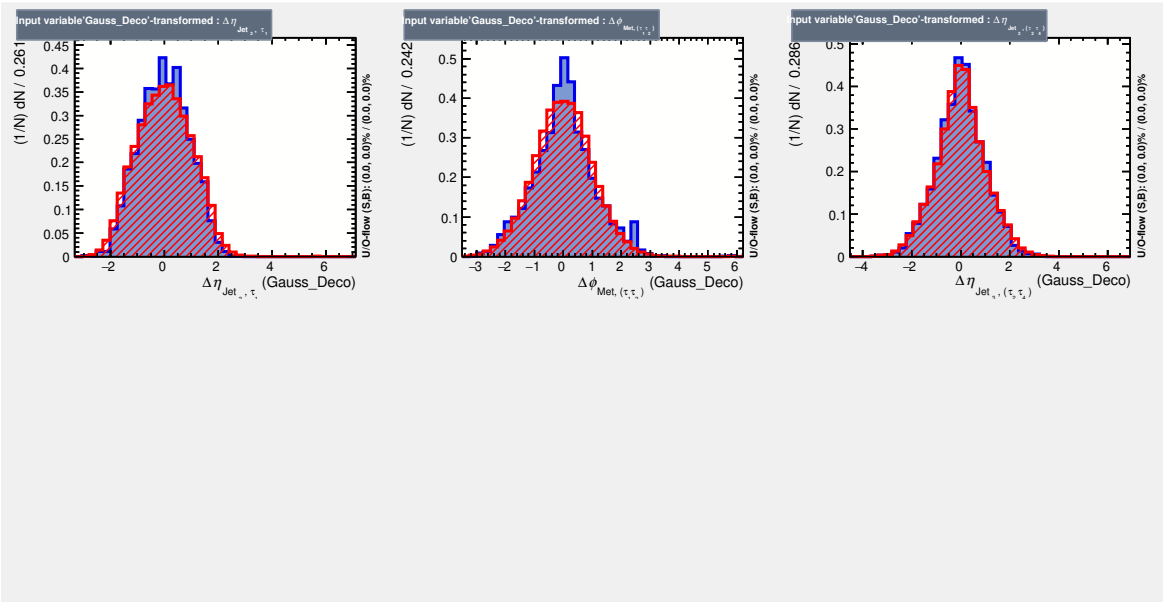


Figure B.7: Further input variables for the Monte Carlo BDT training (7/7).

QCD Multijet BDT

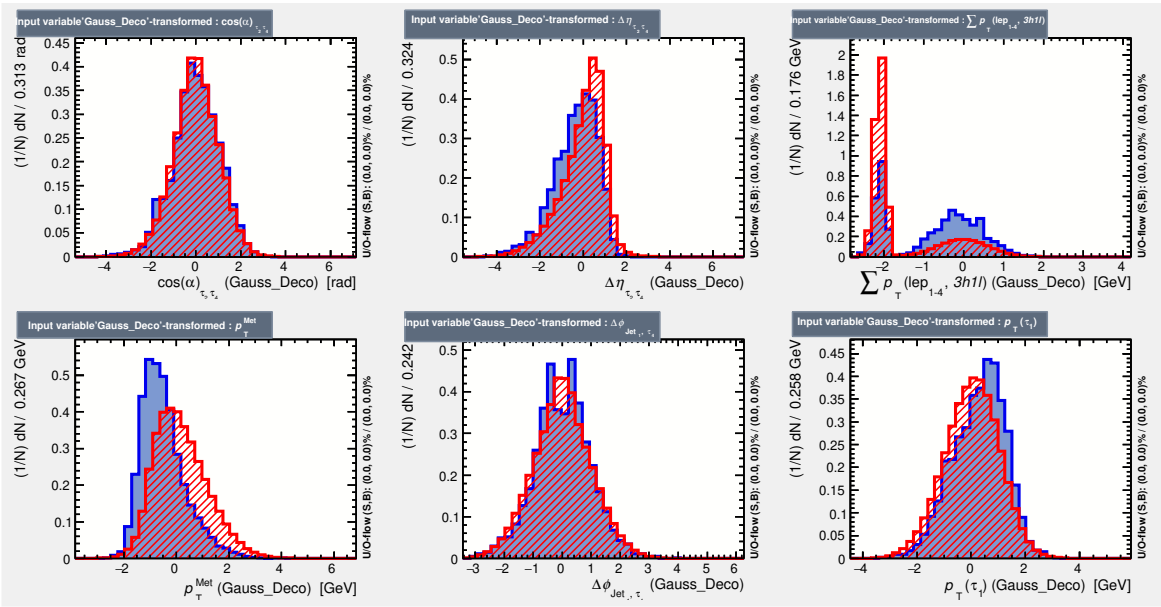


Figure B.8: Further input variables for the Multijet BDT training (1/5).

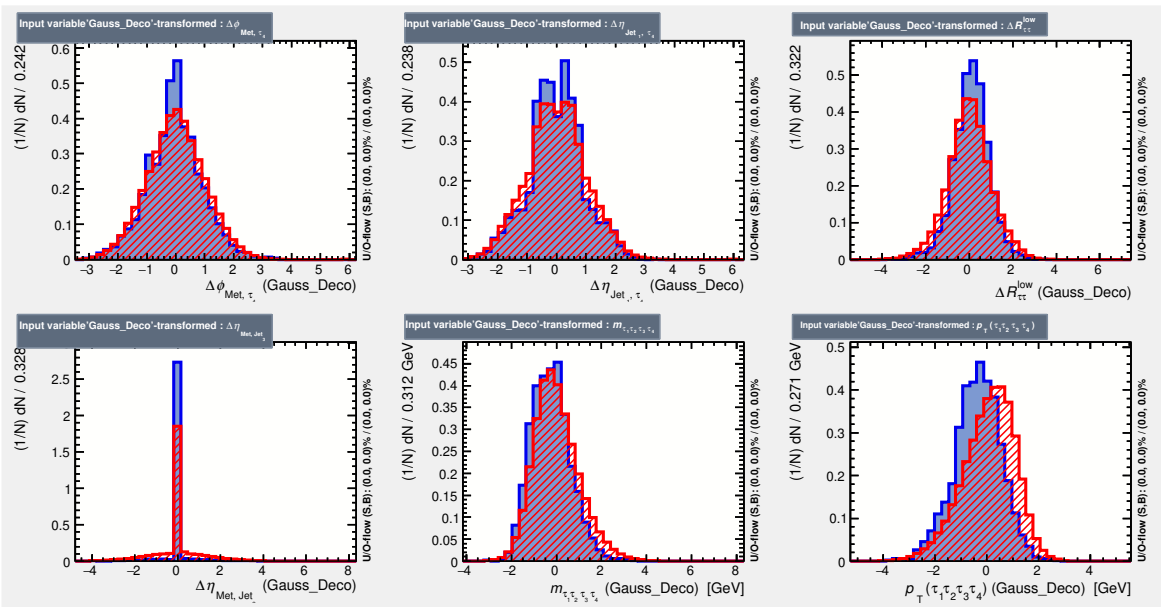


Figure B.9: Further input variables for the Multijet BDT training (2/5).

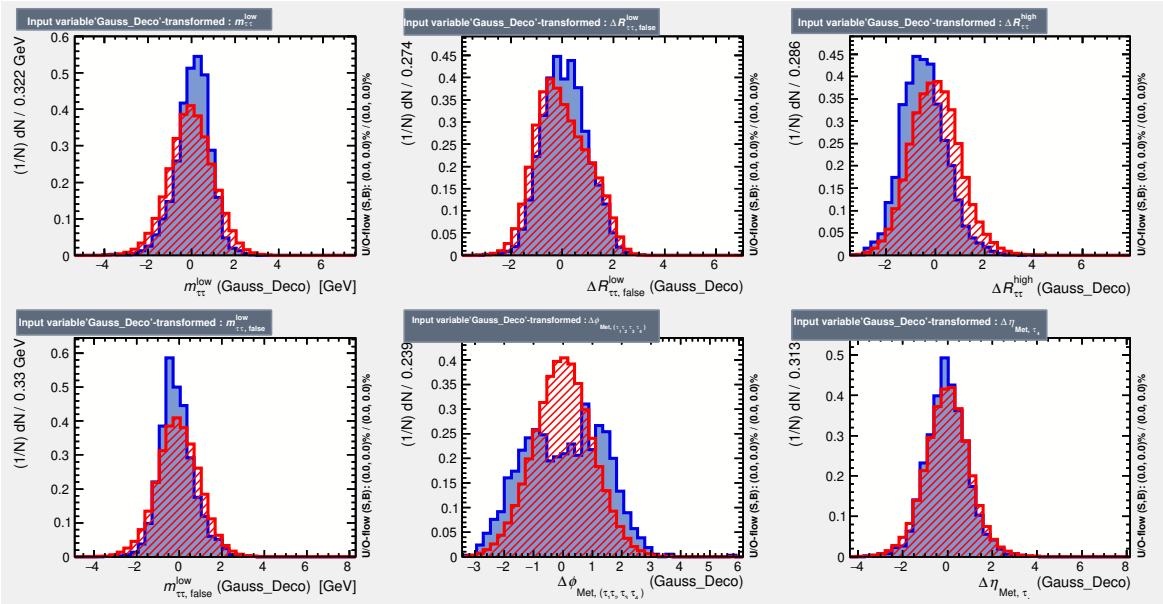


Figure B.10: Further input variables for the Multijet BDT training (3/5).

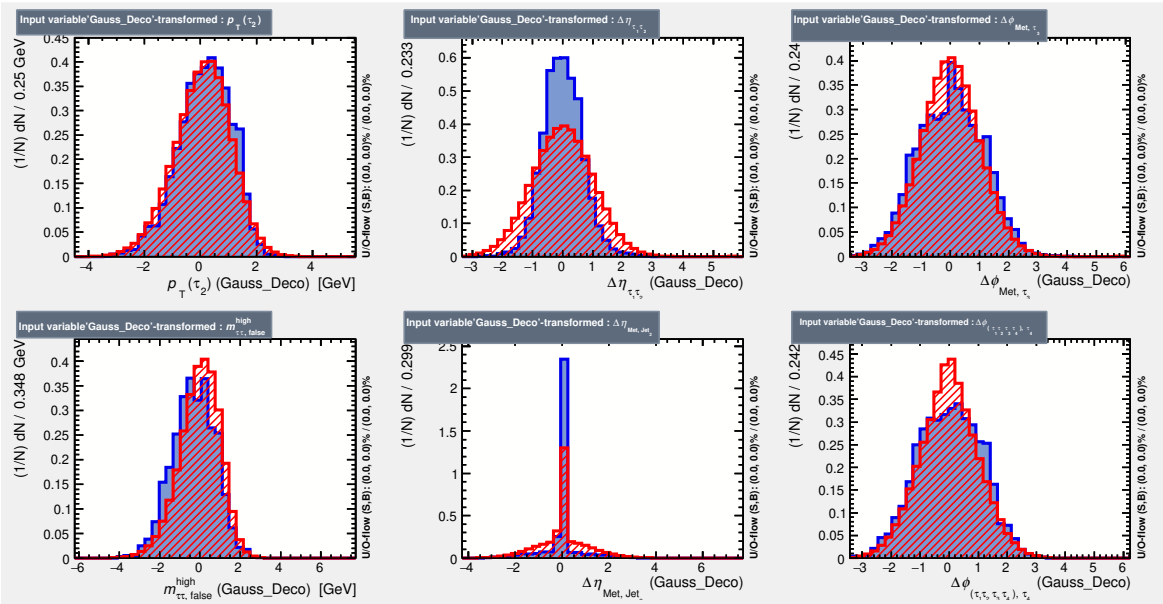


Figure B.11: Further input variables for the Multijet BDT training (4/5).

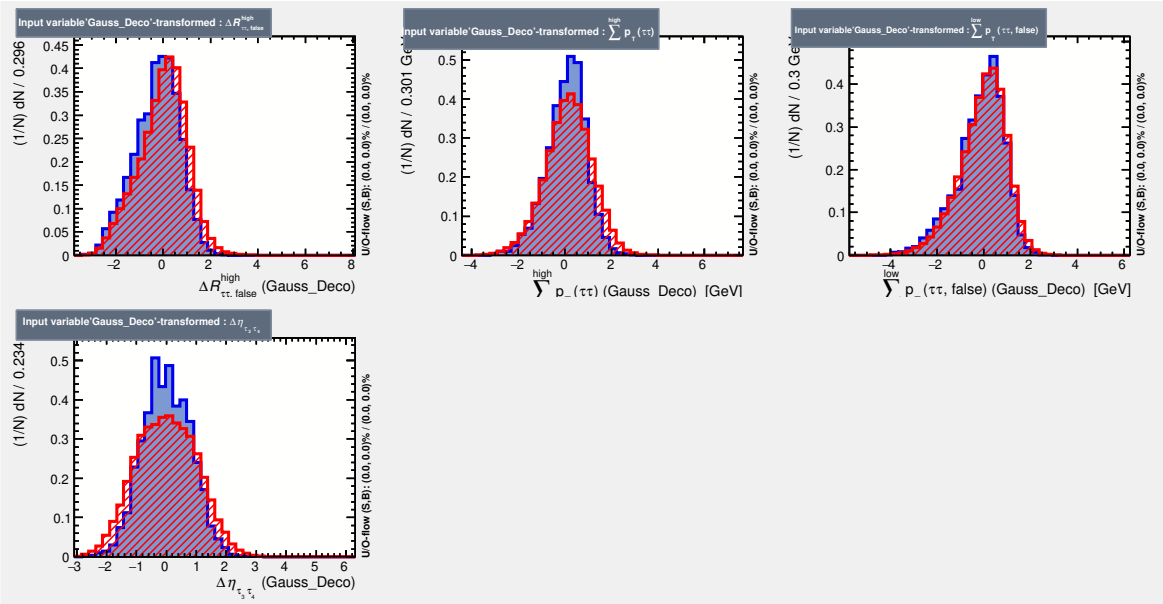


Figure B.12: Further input variables for the Multijet BDT training (5/5).

APPENDIX C

Full Ranking Tables of the Input Variables for the BDT Training

Monte Carlo BDT

Rank	Variable	Importance
1	$\sum p_{\mathrm{T}}(\mathrm{lep}_{1-4}, 3h1l)$	3.410×10^{-2}
2	$p_{\mathrm{T}}^{\mathrm{Met}}$	3.169×10^{-2}
3	$\Delta\eta_{\mathrm{Met,\ Jet_2}}$	3.025×10^{-2}
4	$\Delta\phi_{\mathrm{min}}((\tau_2\tau_4), \mathrm{Met})$	2.863×10^{-2}
5	$\Delta R_{\tau\tau}^{\mathrm{high}}$	2.735×10^{-2}
6	$\Delta\phi_{\tau_1\tau_3}$	2.595×10^{-2}
7	$\Delta\phi_{\mathrm{Met},\tau_4}$	2.567×10^{-2}
8	$m_{\tau\tau}^{\mathrm{low}}$	2.548×10^{-2}
9	$\Delta\phi_{\mathrm{min}}((\tau_1\tau_3), \mathrm{Met})$	2.542×10^{-2}
10	Met significance	2.510×10^{-2}
11	$p_{\mathrm{T}}(\tau_4)$	2.478×10^{-2}
12	$\Delta R_{\tau\tau,\mathrm{false}}^{\mathrm{low}}$	2.400×10^{-2}
13	$\Delta\phi_{\mathrm{Jet}_1,\tau_4}$	2.399×10^{-2}
14	$\Delta\eta_{\tau_1\tau_2}$	2.352×10^{-2}
15	$m_{\tau\tau,\mathrm{false}}^{\mathrm{low}}$	2.334×10^{-2}
16	$p_{\mathrm{T}}(\tau_1)$	2.301×10^{-2}
17	$\Delta\phi_{\mathrm{min}}((\tau_1\tau_2), \mathrm{Met})$	2.298×10^{-2}
18	$\Delta\phi_{\mathrm{Met},(\tau_1\tau_2)}$	2.284×10^{-2}
19	$\Delta\eta_{\tau_1\tau_3}$	2.266×10^{-2}
20	$\Delta R_{\tau\tau}^{\mathrm{low}}$	2.244×10^{-2}

Rank	Variable	Importance
1	$\Delta\phi_{\mathrm{Jet}_1,\tau_4}$	2.908×10^{-2}
2	$\Delta\phi_{(\tau_1\tau_3),(\tau_2\tau_4)}$	2.856×10^{-2}
3	$\Delta\phi_{\mathrm{Jet}_1,\tau_3}$	2.754×10^{-2}
4	$\Delta R_{\tau\tau}^{\mathrm{low}}$	2.678×10^{-2}
5	$\Delta\phi_{(\tau_1\tau_2\tau_3\tau_4),\tau_4}$	2.663×10^{-2}
6	$\Delta\eta_{\mathrm{Met},\tau_4}$	2.640×10^{-2}
7	$\Delta\phi_{\tau_1\tau_3}$	2.626×10^{-2}
8	$\Delta\eta_{\tau_3\tau_4}$	2.580×10^{-2}
9	$\Delta\phi_{\mathrm{Met},\tau_4}$	2.578×10^{-2}
10	$\Delta\eta_{\mathrm{Jet}_1,\tau_3}$	2.558×10^{-2}
11	$\Delta\phi_{(\tau_3\tau_4),(\tau_2\tau_4)}$	2.512×10^{-2}
12	$\Delta\eta_{\tau_1\tau_3}$	2.503×10^{-2}
13	$\Delta R_{\tau_3\tau_4}$	2.467×10^{-2}
14	$\cos(\alpha)_{\tau_2\tau_4}$	2.452×10^{-2}
15	$\Delta\eta_{\tau_2\tau_4}$	2.393×10^{-2}
16	$\Delta\phi_{\mathrm{min}}((\tau_1\tau_2), \mathrm{Met})$	2.369×10^{-2}
17	$\Delta\eta_{\tau_1\tau_2}$	2.347×10^{-2}
18	$\Delta\phi_{\mathrm{min}}((\tau_2\tau_4), \mathrm{Met})$	2.340×10^{-2}
19	Met significance	2.322×10^{-2}
20	$\Delta\eta_{\tau_3\tau_4}$	2.320×10^{-2}

21	$\Delta\phi_{(\tau_1\tau_2\tau_3\tau_4),\tau_4}$	2.243×10^{-2}
22	$\cos(\alpha)_{\tau_2\tau_4}$	2.196×10^{-2}
23	$\Delta\phi_{\min}((\tau_3\tau_4), \mathrm{Met})$	2.186×10^{-2}
24	$\Delta\eta_{\tau_2\tau_4}$	2.166×10^{-2}
25	$\sum \cos\big(\Delta\phi((\tau_1\tau_2), \mathrm{Met})\big)$	2.166×10^{-2}
26	$\Delta R_{\tau_1\tau_2\tau_3\tau_4}$	2.163×10^{-2}
27	$\Delta\eta_{\mathrm{Jet}_2,\tau_1}$	2.085×10^{-2}
28	$\Delta\eta_{\mathrm{Met},\mathrm{Jet}_3}$	2.076×10^{-2}
29	$\Delta\eta_{\tau_3\tau_4}$	2.074×10^{-2}
30	$\Delta\eta_{\tau_3\tau_4}$	2.019×10^{-2}
31	$\Delta\eta_{\mathrm{Jet}_1,\tau_3}$	1.997×10^{-2}
32	$\Delta\eta_{\mathrm{Met},\tau_4}$	1.974×10^{-2}
33	$\Delta\phi_{(\tau_1\tau_3),(\tau_2\tau_4)}$	1.973×10^{-2}
34	$\Delta\phi_{\mathrm{Jet}_2,\mathrm{Jet}_3}$	1.942×10^{-2}
35	$\sum \cos\big(\Delta\phi((\tau_3\tau_4), \mathrm{Met})\big)$	1.938×10^{-2}
36	$\Delta R_{\tau_3\tau_4}$	1.908×10^{-2}
37	$\cos(\alpha)_{\tau_1\tau_3}$	1.897×10^{-2}
38	$\sum \cos\big(\Delta\phi((\tau_1\tau_3), \mathrm{Met})\big)$	1.879×10^{-2}
39	$\Delta\phi_{(\tau_3\tau_4),(\tau_2\tau_4)}$	1.867×10^{-2}
40	$\Delta\eta_{\mathrm{Jet}_1,\tau_3}$	1.856×10^{-2}
41	$\Delta\eta_{\mathrm{Jet}_3,(\tau_2\tau_4)}$	1.824×10^{-2}
42	$\cos(\alpha)_{\tau_3\tau_4}$	1.760×10^{-2}
43	$\Delta R^{\mathrm{high}}_{\tau\tau,\mathrm{false}}$	1.639×10^{-2}
44	$m^{\mathrm{high}}_{\tau\tau,\mathrm{false}}$	1.503×10^{-2}
45	$m_{\tau_1\tau_3}$	1.348×10^{-2}

(a) BDT (MC)

21	$p_{\mathrm{T}}^{\mathrm{Met}}$	2.290×10^{-2}
22	$\cos(\alpha)_{\tau_3\tau_4}$	2.285×10^{-2}
23	$\Delta R^{\mathrm{high}}_{\tau\tau,\mathrm{false}}$	2.282×10^{-2}
24	$\Delta\phi_{\min}((\tau_3\tau_4), \mathrm{Met})$	2.278×10^{-2}
25	$\Delta R^{\mathrm{high}}_{\tau\tau}$	2.257×10^{-2}
26	$\Delta\phi_{\mathrm{Met},(\tau_1\tau_2)}$	2.232×10^{-2}
27	$\Delta R^{\mathrm{low}}_{\tau\tau,\mathrm{false}}$	2.216×10^{-2}
28	$\sum \cos\big(\Delta\phi((\tau_3\tau_4), \mathrm{Met})\big)$	2.169×10^{-2}
29	$\cos(\alpha)_{\tau_1\tau_3}$	2.162×10^{-2}
30	$\Delta\eta_{\mathrm{Met},\mathrm{Jet}_2}$	2.135×10^{-2}
31	$\Delta\eta_{\mathrm{Jet}_3,(\tau_2\tau_4)}$	2.120×10^{-2}
32	$\Delta\phi_{\min}((\tau_1\tau_3), \mathrm{Met})$	2.108×10^{-2}
33	$\Delta R_{\tau_1\tau_2\tau_3\tau_4}$	2.106×10^{-2}
34	$\sum \cos\big(\Delta\phi((\tau_1\tau_2), \mathrm{Met})\big)$	2.104×10^{-2}
35	$\Delta\eta_{\mathrm{Jet}_2,\tau_1}$	2.081×10^{-2}
36	$p_{\mathrm{T}}(\tau_1)$	2.052×10^{-2}
37	$m^{\mathrm{low}}_{\tau\tau}$	1.983×10^{-2}
38	$\sum p_{\mathrm{T}}(\mathrm{lep}_{1-4}, 3h1l)$	1.901×10^{-2}
39	$p_{\mathrm{T}}(\tau_4)$	1.845×10^{-2}
40	$\sum \cos\big(\Delta\phi((\tau_1\tau_3), \mathrm{Met})\big)$	1.833×10^{-2}
41	$\Delta\phi_{\mathrm{Jet}_2,\mathrm{Jet}_3}$	1.682×10^{-2}
42	$\Delta\eta_{\mathrm{Met},\mathrm{Jet}_3}$	1.533×10^{-2}
43	$m^{\mathrm{low}}_{\tau\tau,\mathrm{false}}$	1.450×10^{-2}
44	$m_{\tau_1\tau_3}$	1.187×10^{-2}
45	$m^{\mathrm{high}}_{\tau\tau,\mathrm{false}}$	8.422×10^{-3}

(b) BDTG (MC)

Table C.1: Full classifier-specific rankings of all input variables for the training of the Monte Carlo BDT.

QCD Multijet BDT

Rank	Variable	Importance
1	$p_{\mathrm{T}}^{\mathrm{Met}}$	5.837×10^{-2}
2	Met significance	5.585×10^{-2}
3	$\Delta\phi_{\mathrm{Jet}_1,\tau_4}$	4.507×10^{-2}
4	$\sum p_{\mathrm{T}}(\mathrm{lep}_{1-4}, 3h1l)$	4.455×10^{-2}

Rank	Variable	Importance
1	$p_{\mathrm{T}}^{\mathrm{Met}}$	4.600×10^{-2}
2	Met significance	4.569×10^{-2}
3	$\Delta\eta_{\tau_1\tau_3}$	3.506×10^{-2}
4	$\Delta\phi_{\mathrm{Jet}_1,\tau_4}$	3.457×10^{-2}

5	$\Delta\eta_{\tau_1\tau_3}$	3.535×10^{-2}
6	$\Delta R_{\tau\tau}^{\text{high}}$	3.396×10^{-2}
7	$\Delta\phi_{\text{Met},(\tau_1\tau_2\tau_3\tau_4)}$	3.293×10^{-2}
8	$m_{\tau\tau}^{\text{low}}$	3.284×10^{-2}
9	$p_{\text{T}}(\tau_1\tau_2\tau_3\tau_4)$	3.273×10^{-2}
10	$\Delta\phi_{\text{Met},\tau_4}$	3.246×10^{-2}
11	$\Delta\eta_{\tau_1\tau_2}$	3.154×10^{-2}
12	$\Delta\eta_{\text{Met,Jet}_2}$	3.141×10^{-2}
13	$\Delta\eta_{\text{Met},\tau_4}$	2.978×10^{-2}
14	$\Delta\eta_{\text{Met,Jet}_3}$	2.846×10^{-2}
15	$p_{\text{T}}(\tau_1)$	2.835×10^{-2}
16	$\Delta R_{\tau\tau,\text{false}}^{\text{high}}$	2.772×10^{-2}
17	$m_{\tau\tau,\text{false}}^{\text{low}}$	2.615×10^{-2}
18	$\Delta\eta_{\tau_3\tau_4}$	2.600×10^{-2}
19	$\cos(\alpha)_{\tau_1\tau_2}$	2.593×10^{-2}
20	$\Delta\phi_{(\tau_1\tau_2\tau_3\tau_4),\tau_4}$	2.533×10^{-2}
21	$\Delta R_{\tau\tau}^{\text{low}}$	2.444×10^{-2}
22	$\Delta\eta_{\tau_1\tau_2}$	2.436×10^{-2}
23	$\cos(\alpha)_{\tau_2\tau_4}$	2.374×10^{-2}
24	$\Delta R_{\tau\tau,\text{false}}^{\text{low}}$	2.358×10^{-2}
25	$m_{\tau\tau,\text{false}}^{\text{high}}$	2.334×10^{-2}
26	$\sum^{\text{high}} p_{\text{T}}(\tau\tau)$	2.325×10^{-2}
27	$\Delta\eta_{\text{Jet}_1,\tau_4}$	2.284×10^{-2}
28	$\Delta\eta_{\tau_2\tau_4}$	2.236×10^{-2}
29	$\cos(\alpha)_{\tau_1\tau_3}$	2.203×10^{-2}
30	$m_{\tau_1\tau_2\tau_3\tau_4}$	2.201×10^{-2}
31	$\Delta\phi_{\text{Met},\tau_3}$	2.188×10^{-2}
32	$\Delta R_{\tau_1\tau_2}$	2.187×10^{-2}
33	$p_{\text{T}}(\tau_2)$	2.023×10^{-2}
34	$\sum^{\text{low}} p_{\text{T}}(\tau\tau, \text{false})$	1.930×10^{-2}

(a) BDT (Fakes)

5	$\Delta\eta_{\text{Met},\tau_4}$	3.450×10^{-2}
6	$\Delta R_{\tau\tau,\text{false}}^{\text{low}}$	3.443×10^{-2}
7	$\Delta\eta_{\tau_3\tau_4}$	3.389×10^{-2}
8	$\cos(\alpha)_{\tau_1\tau_3}$	3.378×10^{-2}
9	$\Delta\phi_{\text{Met},\tau_4}$	3.368×10^{-2}
10	$\Delta R_{\tau\tau}^{\text{high}}$	3.285×10^{-2}
11	$\Delta\eta_{\tau_2\tau_4}$	3.274×10^{-2}
12	$\cos(\alpha)_{\tau_1\tau_2}$	3.274×10^{-2}
13	$\Delta\phi_{(\tau_1\tau_2\tau_3\tau_4),\tau_4}$	3.272×10^{-2}
14	$\cos(\alpha)_{\tau_2\tau_4}$	3.240×10^{-2}
15	$\Delta R_{\tau_1\tau_2}$	3.185×10^{-2}
16	$\Delta\phi_{\text{Met},(\tau_1\tau_2\tau_3\tau_4)}$	3.163×10^{-2}
17	$\Delta\phi_{\text{Met},\tau_3}$	3.141×10^{-2}
18	$\Delta\eta_{\tau_1\tau_2}$	3.122×10^{-2}
19	$m_{\tau\tau}^{\text{low}}$	2.914×10^{-2}
20	$\Delta\eta_{\text{Jet}_1,\tau_4}$	2.894×10^{-2}
21	$\Delta\eta_{\tau_1\tau_2}$	2.886×10^{-2}
22	$\Delta R_{\tau\tau}^{\text{low}}$	2.846×10^{-2}
23	$\sum p_{\text{T}}(\text{lep}_{1-4}, 3h1l)$	2.802×10^{-2}
24	$\Delta\eta_{\text{Met,Jet}_2}$	2.781×10^{-2}
25	$p_{\text{T}}(\tau_1\tau_2\tau_3\tau_4)$	2.546×10^{-2}
26	$p_{\text{T}}(\tau_1)$	2.388×10^{-2}
27	$\Delta R_{\tau\tau,\text{false}}^{\text{high}}$	2.268×10^{-2}
28	$\sum^{\text{high}} p_{\text{T}}(\tau\tau)$	2.236×10^{-2}
29	$\Delta\eta_{\text{Met,Jet}_3}$	2.214×10^{-2}
30	$m_{\tau_1\tau_2\tau_3\tau_4}$	2.193×10^{-2}
31	$\sum^{\text{low}} p_{\text{T}}(\tau\tau, \text{false})$	1.907×10^{-2}
32	$m_{\tau\tau,\text{false}}^{\text{high}}$	1.824×10^{-2}
33	$m_{\tau\tau,\text{false}}^{\text{low}}$	1.812×10^{-2}
34	$p_{\text{T}}(\tau_2)$	1.376×10^{-2}

(b) BDTG (Fakes)

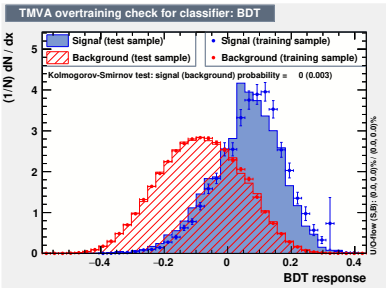
Table C.2: Full classifier-specific rankings of all input variables for the training of the QCD Multijet BDT.

APPENDIX D

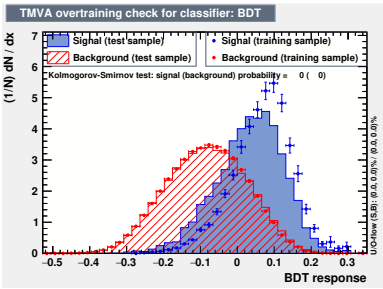
BDT Overtraining Results for Different Hyperparameter Setups

In this chapter, the overtraining results of the trained BDT classifiers for different hyperparameter setups are shown. The final selection of a tree depth of 3 and a minimum node size of 5 % (cf. tab. 7.2) has been made based on the plots for the *BDTG (MC)* and *BDTG (Fakes)* classifiers in combination with table 7.7 and 7.8 and is highlighted in the captions below.

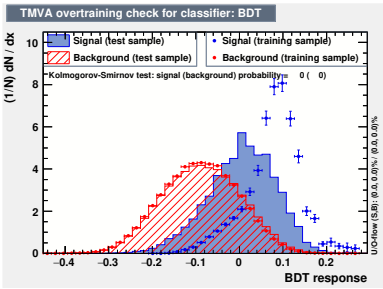
Monte Carlo BDT



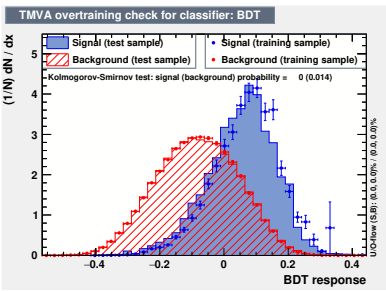
(a) Depth 2, min. node size 2 %



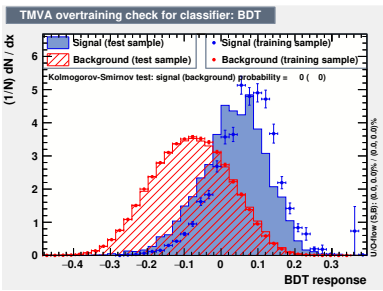
(b) Depth 3, min. node size 2 %



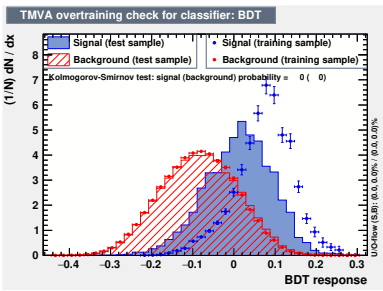
(c) Depth 5, min. node size 2 %



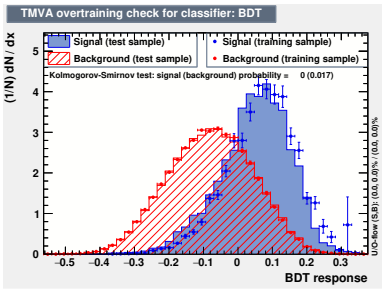
(d) Depth 2, min. node size 5 %



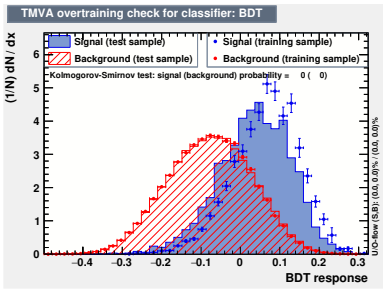
(e) **Depth 3, min. node size 5 %**



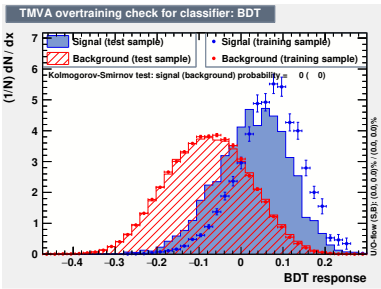
(f) Depth 5, min. node size 5 %



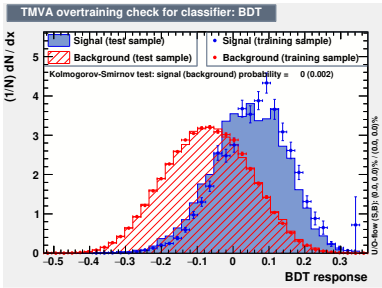
(g) Depth 2, min. node size 10 %



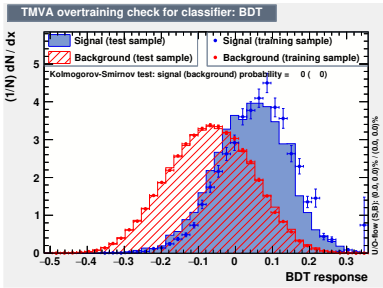
(h) Depth 3, min. node size 10 %



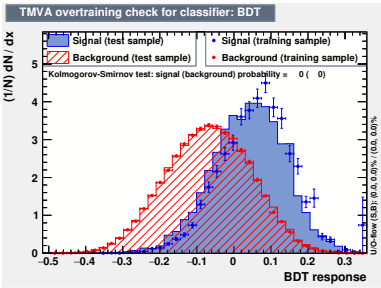
(i) Depth 5, min. node size 10 %



(j) Depth 2, min. node size 20 %

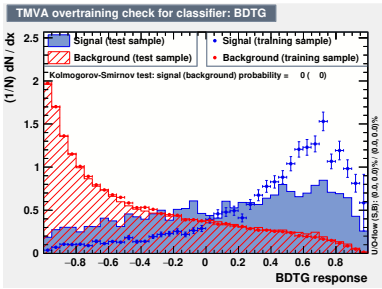


(k) Depth 3, min. node size 20 %

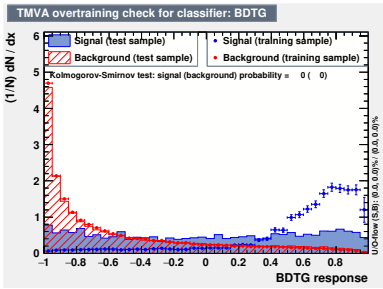


(l) Depth 5, min. node size 20 %

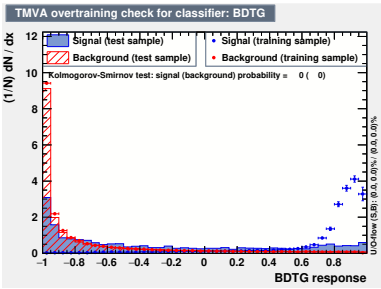
Figure D.1: Overtraining results of the *BDT (MC)* classifier for different hyperparameter setups. The effect is large for high depths and small minimum node sizes and decreases especially when higher values for the second parameter are chosen. On the other hand, this increase also leads to a larger overlap between the signal and background distributions, i.e. to a worse separation performance.



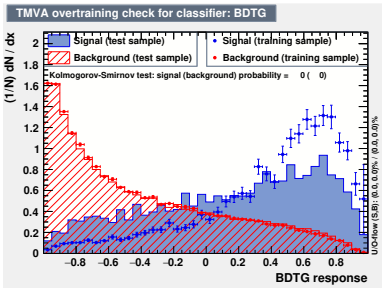
(a) Depth 2, min. node size 2 %



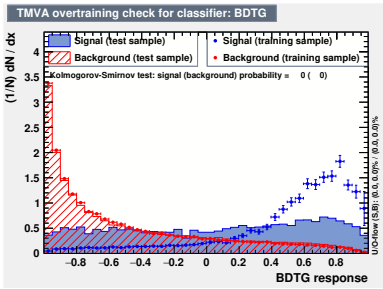
(b) Depth 3, min. node size 2 %



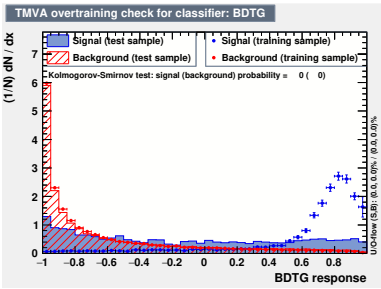
(c) Depth 5, min. node size 2 %



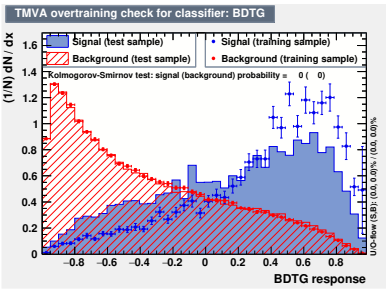
(d) Depth 2, min. node size 5 %



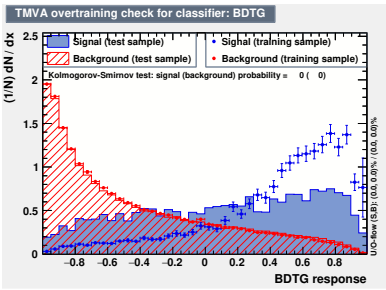
(e) **Depth 3, min. node size 5 %**



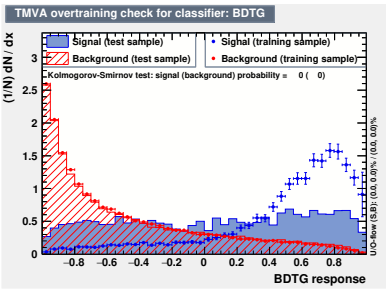
(f) Depth 5, min. node size 5 %



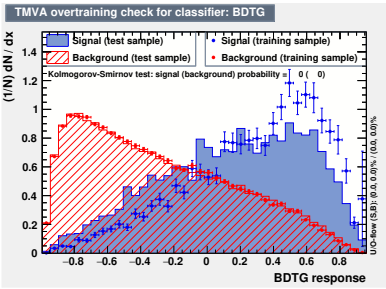
(g) Depth 2, min. node size 10 %



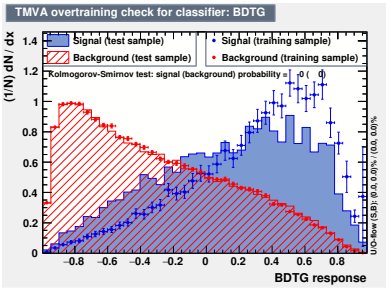
(h) Depth 3, min. node size 10 %



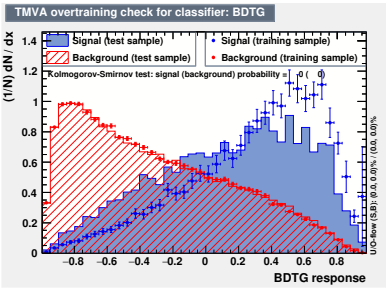
(i) Depth 5, min. node size 10 %



(j) Depth 2, min. node size 20 %



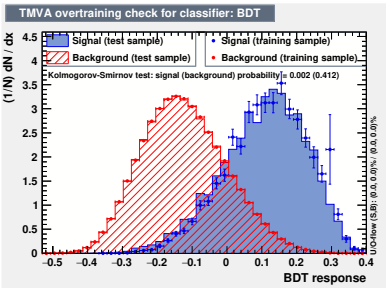
(k) Depth 3, min. node size 20 %



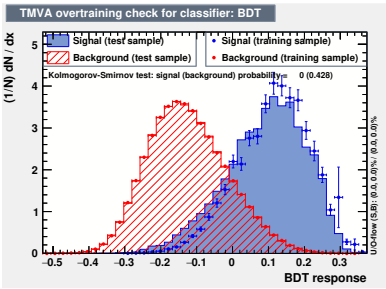
(l) Depth 5, min. node size 20 %

Figure D.2: Overtraining results of the *BDTG (MC)* classifier for different hyperparameter setups. The effect is large for high depths and small minimum node sizes and decreases especially when higher values for the second parameter are chosen. On the other hand, this increase also leads to a larger overlap between the signal and background distributions, i.e. to a worse separation performance.

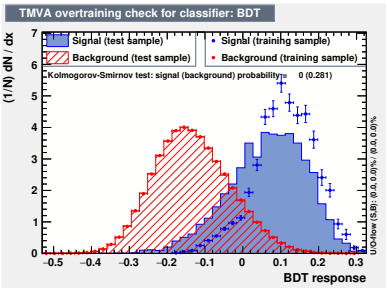
QCD Multijet BDT



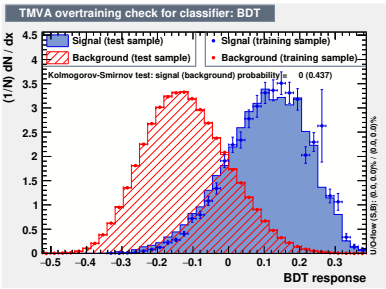
(a) Depth 2, min. node size 2 %



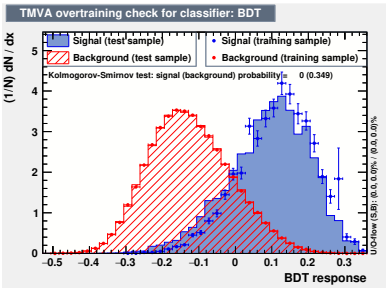
(b) Depth 3, min. node size 2 %



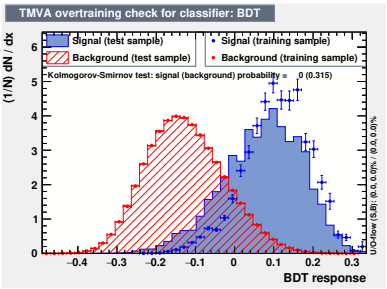
(c) Depth 5, min. node size 2 %



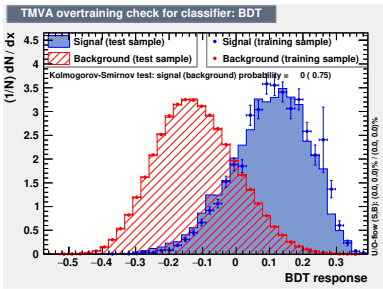
(d) Depth 2, min. node size 5 %



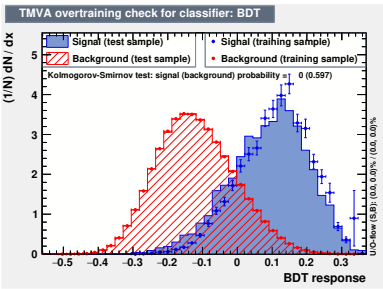
(e) **Depth 3, min. node size 5 %**



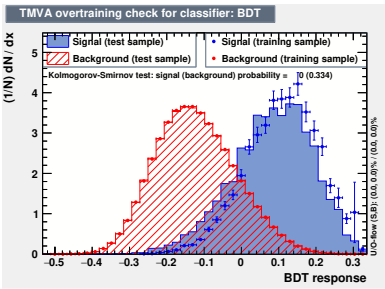
(f) Depth 5, min. node size 5 %



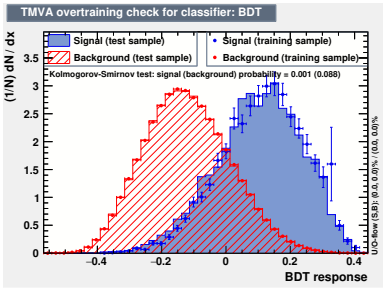
(g) Depth 2, min. node size 10 %



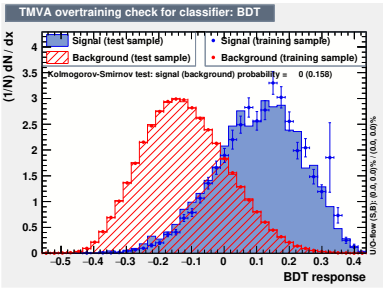
(h) Depth 3, min. node size 10 %



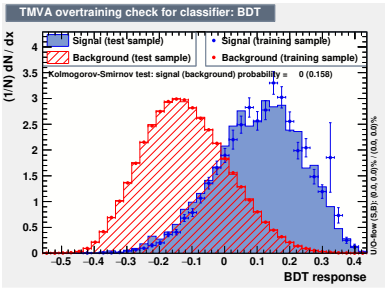
(i) Depth 5, min. node size 10 %



(j) Depth 2, min. node size 20 %

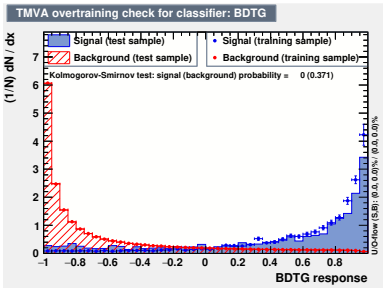


(k) Depth 3, min. node size 20 %

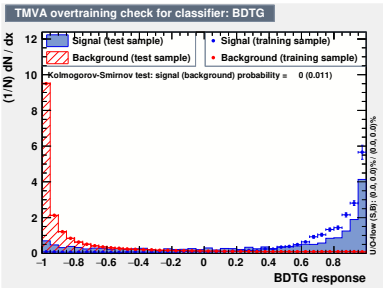


(l) Depth 5, min. node size 20 %

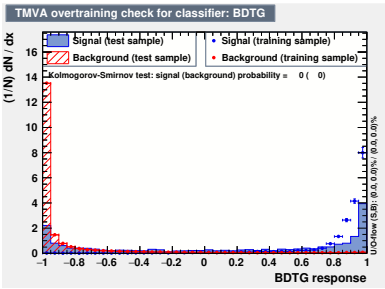
Figure D.3: Overtraining results of the *BDT (Fakes)* classifier for different hyperparameter setups. The effect is large for high depths and small minimum node sizes and decreases especially when higher values for the second parameter are chosen. On the other hand, this increase also leads to a larger overlap between the signal and background distributions, i.e. to a worse separation performance.



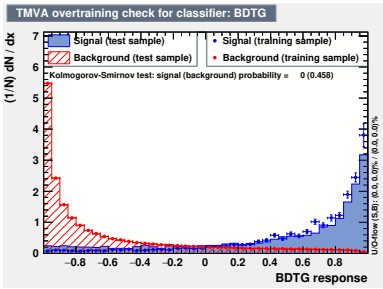
(a) Depth 2, min. node size 2 %



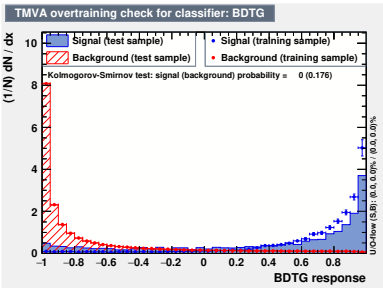
(b) Depth 3, min. node size 2 %



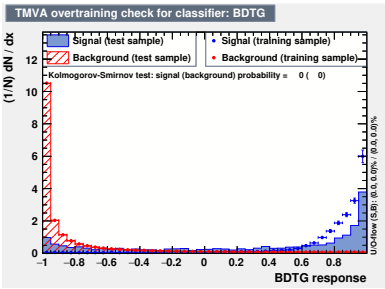
(c) Depth 5, min. node size 2 %



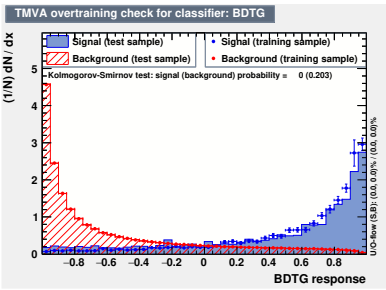
(d) Depth 2, min. node size 5 %



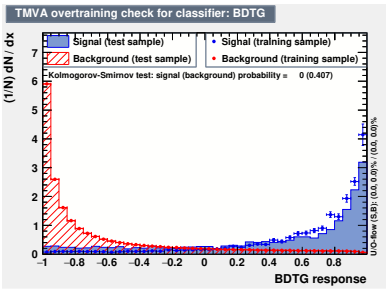
(e) **Depth 3, min. node size 5 %**



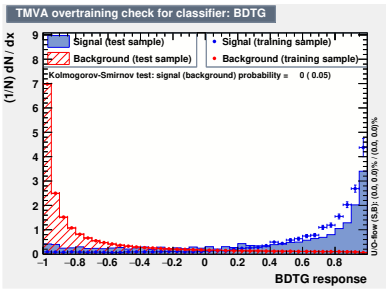
(f) Depth 5, min. node size 5 %



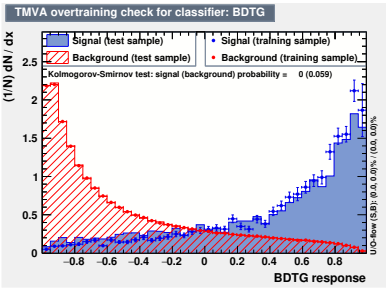
(g) Depth 2, min. node size 10 %



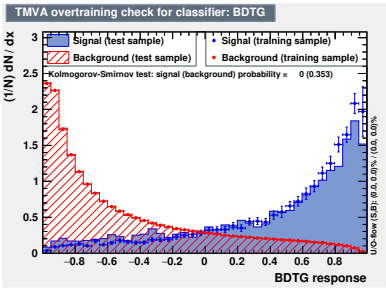
(h) Depth 3, min. node size 10 %



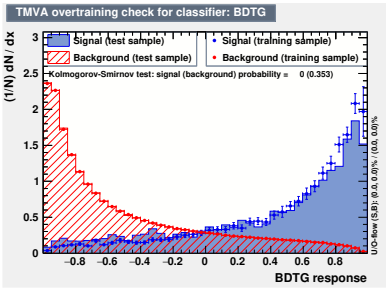
(i) Depth 5, min. node size 10 %



(j) Depth 2, min. node size 20 %



(k) Depth 3, min. node size 20 %



(l) Depth 5, min. node size 20 %

Figure D.4: Overtraining results of the *BDTG (Fakes)* classifier for different hyperparameter setups. The effect is large for high depths and small minimum node sizes and decreases especially when higher values for the second parameter are chosen. On the other hand, this increase also leads to a larger overlap between the signal and background distributions, i.e. to a worse separation performance.

APPENDIX E

Pre-Fit Event Yield Tables

The pre-fit events yields in the different fit regions constructed in section 7.3 are given below.

Process	Bin									
	1	2	3	4	5	6	7	8	9	10
$ZZ \to 4\tau$	0.002	0	0.008	0.040	0.051	0.118	0.109	0.130	0.041	0.082
V + Jets	0	0	0	0	0	0	0	0.050	0	0
Higgs	0	0	0	0	0.001	0.001	0.007	0.011	0.037	0.027
$Z \to \tau\tau$	0	0	0	0	0	0	0	0.182	0.022	0
Multiboson	0	0	0	0	0.002	0.003	0.007	0.032	0.015	0.101
$t\bar{t}$	0	0	0	0	0.509	1.027	1.195	0.173	0.338	0.086
Top	0	0	0	0	0	0	0	0	0	0
Fakes (QCD)	3850.972	5047.551	32962.807	46902.751	23847.177	8778.283	2425.485	569.188	90.695	12.935
Fakes (MC)	169.028	237.449	1448.193	21575.249	12240.630	8292.301	3294.687	962.203	214.729	30.044
$\sum$ MC	0	0	0	0	0.512	1.031	1.209	0.448	0.412	0.214
$\sum$ Fakes	4020.000	5285.000	34411.000	68478.000	36087.807	17070.584	5720.172	1531.391	305.424	42.979
Total	4020 ± 608	5285 ± 798	34411 ± 5167	68478 ± 10277	36088 ± 5418	17072 ± 2566	5721 ± 863	1532 ± 234	306 ± 50	43 ± 10
Data	3779	5266	34430	67433	35485	16679	5409	1476	319	51

Table E.1: Bin-separated pre-fit event yields in the $CR_{MCLowFLow}$ region.

Process	Bin									
	1	2	3	4	5	6	7	8	9	10
$ZZ \to 4\tau$	0.009	0.002	0.039	0.149	0.213	0.453	0.447	0.626	0.547	0.538
V + Jets	0	0	0	0	0	0	0	0	0	0
Higgs	0	0	0	0	0.006	0.001	0.019	0.049	0.088	0.212
$Z \to \tau\tau$	0	0	0	0	0	0	0.088	0.449	0.051	0
Multiboson	0	0	0	0	0	0.001	0.001	0	0.051	0.205
$t\bar{t}$	0	0	0	0	0.081	0.424	0.597	0.258	0.268	0.084
Top	0	0	0	0	0	0	0	0	0	0
Fakes (QCD)	7820.945	10732.229	80830.385	110436.935	78192.898	35482.149	11531.468	2586.003	359.354	43.161
Fakes (MC)	371.055	535.771	4167.615	16969.805	11037.932	8555.766	3335.659	1299.646	298.488	71.628
$\sum$ MC	0	0	0	0	0.087	0.426	0.705	0.756	0.458	0.501
$\sum$ Fakes	8192.000	11268.000	84998.000	127406.740	89230.830	44037.915	14867.127	3885.650	657.842	114.789
Total	8192 ± 809	11268 ± 1100	84998 ± 8120	127407 ± 11202	89231 ± 7925	44039 ± 3682	14868 ± 1235	3887 ± 332	659 ± 86	116 ± 34
Data	8077	11171	85626	127480	88196	44690	15056	3987	711	106

Table E.2: Bin-separated pre-fit event yields in the $CR_{MCHighFLow}$ region.

Process	Bin									
	1	2	3	4	5	6	7	8	9	10
$ZZ \to 4\tau$	0	0.001	0.003	0.014	0.013	0.032	0.039	0.039	0.048	0.047
V + Jets	0	0	0	0	0	0	0	0	0	0
Higgs	0	0	0	0	0.002	0.003	0.017	0.050	0.100	0.358
$Z \to \tau\tau$	0	0	0	0	0	0	0	0.080	0.079	0
Multiboson	0	0	0	0	0	0	0.031	0.001	0.022	0.024
$t\bar{t}$	0	0	0	0	0	0.787	1.287	0.747	0.862	0.170
Top	0	0	0	0	0	0	0	0.313	0	0
Fakes (QCD)	292.194	292.195	1842.462	1179.211	623.609	166.708	75.083	5.546	0	1.909
Fakes (MC)	97.806	143.805	860.538	24311.789	11488.391	5902.771	2066.258	631.839	117.917	20.066
$\sum$ MC	0	0	0	0	0.002	0.790	1.335	1.191	1.063	0.552
$\sum$ Fakes	390.000	436.000	2703.000	25491.000	12112.000	6069.479	2141.341	637.385	117.917	21.975
Total	390 ± 53	436 ± 60	2703 ± 236	25491 ± 2454	12112 ± 1171	6070 ± 610	2143 ± 225	639 ± 80	119 ± 24	23 ± 10
Data	379	495	2759	25833	12202	6360	2183	634	152	34

Table E.3: Bin-separated pre-fit event yields in the CR_{Top} region.

Process	Bin									
	1	2	3	4	5	6	7	8	9	10
$ZZ \to 4\tau$	0	0	0.011	0.046	0.125	0.288	0.516	0.783	0.684	1.236
V + Jets	0	0	0	0	0	0	0	0	0	0
Higgs	0	0	0	0	0.001	0.006	0.007	0.061	0.115	0.641
$Z \to \tau\tau$	0	0	0	0	0	0	0	0	0.063	0
Multiboson	0	0	0	0	0	0.002	0.001	0.016	0.049	0.075
$t\bar{t}$	0	0	0	0	0.086	0.250	0.924	0.244	0.081	0
Top	0	0	0	0	0	0	0	0	0	0
Fakes (QCD)	207.970	260.854	2405.562	4266.748	4121.840	2196.596	758.230	173.315	36.374	2.330
Fakes (MC)	24.030	55.146	535.438	6884.330	4575.986	3202.943	1254.253	482.512	98.599	24.641
$\sum$ MC	0	0	0	0	0.087	0.258	0.932	0.321	0.308	0.716
$\sum$ Fakes	232.000	316.000	2941.000	11151.078	8697.826	5399.539	2012.483	655.827	134.973	26.971
Total	232 ± 44	316 ± 48	2941 ± 276	11151 ± 848	8698 ± 649	5400 ± 429	2014 ± 191	657 ± 91	136 ± 33	29 ± 16
Data	246	345	3220	11829	9058	5533	2159	723	167	20

Table E.4: Bin-separated pre-fit event yields in the $SR_{MCHighFMed}$ region.

Process	Bin									
	1	2	3	4	5	6	7	8	9	10
$ZZ \to 4\tau$	0	0	0.005	0.007	0.081	0.185	0.482	0.784	1.164	3.279
V + Jets	0	0	0	0	0	0	0	0	0	0
Higgs	0	0	0	0	0	0.001	0.029	0.512	0.203	1.517
$Z \to \tau\tau$	0	0	0	0	0	0	0	0	0.064	0.063
Multiboson	0	0	0	0	0	0	0.002	0.008	0.037	0.084
$t\bar{t}$	0	0	0	0.250	0.084	0.768	0.255	0.417	0	0
Top	0	0	0	0	0	0	0	0	0	0
Fakes (QCD)	29.921	43.093	404.776	682.526	838.316	390.646	116.552	26.008	8.974	3.572
Fakes (MC)	9.079	14.907	181.224	3494.252	2169.684	1699.982	771.136	265.566	76.505	12.404
$\sum$ MC	0	0	0	0.250	0.084	0.769	0.286	0.937	0.304	1.664
$\sum$ Fakes	39.000	58.000	586.000	4176.778	3008.000	2090.628	887.688	291.574	85.479	15.976
Total	39 ± 13	58 ± 18	586 ± 70	4177 ± 379	3008 ± 261	2092 ± 212	888 ± 129	293 ± 71	87 ± 37	21 ± 21
Data	41	70	636	4585	3253	2366	967	372	73	16

Table E.5: Bin-separated pre-fit event yields in the $SR_{MCHighFHigh}$ region.

Process	Bin									
	1	2	3	4	5	6	7	8	9	10
$ZZ \to 4\tau$	0	0	0.002	0.004	0.013	0.030	0.017	0.075	0.035	0.078
V + Jets	0	0	0	0	0	0	0	0	0	0
Higgs	0	0	0	0.001	0.001	0	0.001	0.002	0.016	0.036
$Z \to \tau\tau$	0	0	0	0	0	0	0	0	0	0
Multiboson	0	0	0	0	0	0	0	0	0.001	0.062
$t\bar{t}$	0	0	0	0	0	0.602	0.348	0.349	0.086	0
Top	0	0	0	0	0	0	0	0	0	0
Fakes (QCD)	17.196	22.233	119.253	95.933	95.381	53.106	19.274	5.669	1.865	0
Fakes (MC)	5.804	6.767	48.747	2601.067	1396.626	902.458	352.726	107.323	16.066	2.998
$\sum$ MC	0	0	0	0.001	0.001	0.602	0.349	0.351	0.103	0.098
$\sum$ Fakes	23.000	29.000	168.000	2697.000	1492.007	955.564	372.000	112.992	17.931	2.998
Total	23 ± 11	29 ± 11	168 ± 30	2697 ± 280	1492 ± 166	956 ± 111	372 ± 57	113 ± 27	18 ± 10	3 ± 5
Data	17	21	156	3044	1784	1105	423	130	21	7

Table E.6: Bin-separated pre-fit event yields in the $VR_{MCLowFMed+}$ region.

APPENDIX F

Post-Fit Event Yield Tables

The post-fit events yields in the different fit regions constructed in section 7.3 are given below. In contrast to the pre-fit yields in appendix E, they can only be given for signal, data and the summed Monte Carlo and fake backgrounds, as the fit also treats these processes as one sample each. Due to the negative value of the fitted signal strength, the signal event yields are negative as well.

Process	Bin									
	1	2	3	4	5	6	7	8	9	10
$ZZ \to 4\tau$	-0.001	0	-0.004	-0.020	-0.025	-0.057	-0.053	-0.063	-0.020	-0.040
MC	0	0	0	0	0.669	1.350	1.582	0.588	0.541	0.279
Fakes	4029.967	5298.104	34496.320	68647.786	36177.284	17112.909	5734.355	1535.188	306.181	43.085
Total	4030 ± 81	5298 ± 93	34496 ± 245	68648 ± 356	36178 ± 246	17114 ± 163	5736 ± 92	1536 ± 47	307 ± 21	43 ± 8
Data	3779	5266	34430	67433	35485	16679	5409	1476	319	51

Table F.1: Bin-separated post-fit event yields in the $CR_{MCLowFLow}$ region.

Process	Bin									
	1	2	3	4	5	6	7	8	9	10
$ZZ \to 4\tau$	-0.004	-0.001	-0.019	-0.073	-0.104	-0.221	-0.218	-0.305	-0.266	-0.262
MC	0	0	0	0	0.114	0.557	0.922	0.986	0.601	0.660
Fakes	8212.311	11295.938	85208.746	127722.636	89452.072	44147.104	14903.989	3895.284	659.473	115.073
Total	8212 ± 118	11296 ± 137	85209 ± 408	127723 ± 515	89452 ± 409	44147 ± 271	14905 ± 150	3896 ± 75	660 ± 31	115 ± 13
Data	8077	11171	85626	127480	88196	44690	15056	3987	711	106

Table F.2: Bin-separated post-fit event yields in the $CR_{MCHighFLow}$ region.

Process	Bin									
	1	2	3	4	5	6	7	8	9	10
$ZZ \to 4\tau$	0	0	-0.002	-0.007	-0.006	-0.016	-0.019	-0.019	-0.024	-0.023
MC	0	0	0	0	0.002	1.035	1.748	1.561	1.392	0.723
Fakes	390.967	437.081	2709.702	25554.204	12142.031	6084.528	2146.651	638.966	118.209	22.029
Total	391 ± 24	437 ± 28	2710 ± 66	25554 ± 208	12142 ± 139	6086 ± 97	2148 ± 56	641 ± 30	120 ± 14	23 ± 8
Data	379	495	2759	25833	12202	6360	2183	634	152	34

Table F.3: Bin-separated post-fit event yields in the CR_{Top} region.

Process	Bin									
	1	2	3	4	5	6	7	8	9	10
$ZZ \to 4\tau$	0	0	-0.005	-0.022	-0.061	-0.140	-0.252	-0.382	-0.333	-0.602
MC	0	0	0	0	0.113	0.339	1.221	0.420	0.403	0.937
Fakes	232.575	316.783	2948.292	11178.727	8719.391	5412.926	2017.473	657.453	135.307	27.037
Total	233 ± 20	317 ± 23	2948 ± 70	11179 ± 134	8719 ± 116	5413 ± 89	2018 ± 54	657 ± 30	135 ± 14	27 ± 7
Data	246	345	3220	11829	9058	5533	2159	723	167	20

Table F.4: Bin-separated post-fit event yields in the $SR_{MCHighFMed}$ region.

Process	Bin									
	1	2	3	4	5	6	7	8	9	10
$ZZ \to 4\tau$	0	0	-0.003	-0.003	-0.040	-0.090	-0.235	-0.382	-0.567	-1.598
MC	0	0	0	0.327	0.110	1.007	0.376	1.226	0.398	2.178
Fakes	39.097	58.144	587.453	4187.134	3015.458	2095.811	889.889	292.297	85.691	16.016
Total	39 ± 8	58 ± 11	587 ± 31	4187 ± 80	3016 ± 68	2097 ± 55	890 ± 35	293 ± 20	86 ± 11	17 ± 8
Data	41	70	636	4585	3253	2366	967	372	73	16

Table F.5: Bin-separated post-fit event yields in the $SR_{MCHighFHigh}$ region.

Process	Bin									
	1	2	3	4	5	6	7	8	9	10
$ZZ \to 4\tau$	0	0	-0.001	-0.002	-0.006	-0.015	-0.008	-0.037	-0.017	-0.038
MC	0	0	0	0.001	0.001	0.789	0.456	0.460	0.134	0.128
Fakes	23.057	29.072	168.417	2703.687	1495.706	957.934	372.922	113.272	17.976	3.005
Total	23 ± 7	29 ± 7	168 ± 16	2704 ± 64	1496 ± 47	959 ± 36	373 ± 22	114 ± 12	18 ± 6	3 ± 4
Data	17	21	156	3044	1784	1105	423	130	21	7

Table F.6: Bin-separated post-fit event yields in the $VR_{MCLowFMed+}$ region.

List of Figures

2.1 The Standard Model of particle physics . . . 4
2.2 Feynman diagrams of the Bhabha scattering process . . . 5
2.3 $e^+e^- \to q\bar{q}$ cross section measurements near the Z resonance . . . 10
2.4 Standard Model total production cross section measurements of different particle states involving gauge bosons and top quarks . . . 10
2.5 Leading-order Feynman diagrams for SM ZZ production in pp-collisions . . . 11
2.6 Example Feynman diagram for BSM ZZ production including a triple gauge coupling 11
2.7 The potential $V(\phi) = \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2$ for a complex scalar field ϕ . . . 12
2.8 Dominant Higgs boson production modes at the LHC . . . 13
2.9 Branching ratios of the Higgs boson depending on its mass M_H . . . 14
2.10 Branching ratios of the τ-lepton into its dominant decay channels . . . 14

3.1 Setup of the CERN accelerator complex . . . 16
3.2 The ATLAS detector . . . 17
3.3 Coordinate system of the ATLAS detector . . . 18
3.4 Overview of the ATLAS inner detector . . . 19
3.5 Overview of the ATLAS calorimeters . . . 20
3.6 Typical two-jet event in the CMS detector resulting from a proton-proton collision . . 22
3.7 Hadron-hadron collision in a particle accelerator simulated by a Monte Carlo event generator . . . 25

4.1 Schematic for training and evaluation of a MVA method with TMVA . . . 28
4.2 Example for a BDT classifier distribution . . . 29
4.3 Sketch of a simple decision tree . . . 30

5.1 Efficiency for reconstruction of all charged tracks stemming from the tau decay for 1-prong and 3-prong taus as a function of $\tau_{\text{had-vis}}$ p_{T} . . . 34
5.2 Distribution of the $f_{\text{track}}^{\text{EM}}$ BDT training variable used for tau identification . . . 36
5.3 Tau identification score obtained from the BDT for tau and jet events . . . 36
5.4 τ^{jetBDT} score distribution of different physics processes for an integrated luminosity of 3 000 fb . . . 38
5.5 The probability value τ^{prob} as function of τ^{jetBDT} for different physics processes . . . 39
5.6 Qualitative separation power of the tau probability product for multi-tau events . . . 39
5.7 Rejection power against fake-tau events as a function of truth-tau signal efficiency for different physics processes and tau identification methods . . . 40

6.1 Possible sub-channels for the $hh \to 4\tau$ process and their branching ratios . . . 43

6.2 Cutflow histogram of the cuts placed on top of the DAOD cuts and the resulting background modeling for two observables of interest . . . 46
6.3 Result of the data driven fake estimation from unphysical charge regions . . . 48
6.4 Distribution of the fake tau background for different approaches of Monte Carlo subtraction in the fake estimation technique . . . 49
6.5 Result of the advanced fake tau estimation from unphysical charge regions . . . 49
6.6 Logarithmic visualization of the results shown in figure 6.5 . . . 50

7.1 Distributions of the different cut variables used to enhance the signal-to-background separation in the $ZZ \rightarrow 4\tau$ channel . . . 52
7.2 Invariant mass distribution of the 4τ system after the application of a b-veto and different cuts on $\log\left(\prod \tau^{\text{prob}}_{\text{i=1-3,4}}\right)$ . . . 53
7.3 Exemplary input variable distributions of the decision tree trained for discrimination of the $ZZ \rightarrow 4\tau$ signal against QCD multijet events . . . 56
7.4 Exemplary input variable distributions of the decision tree trained for discrimination of the $ZZ \rightarrow 4\tau$ signal against Monte Carlo background events . . . 57
7.5 Overtraining check for the Monte Carlo BDT . . . 59
7.6 Overtraining check for the QCD multijet BDT . . . 60
7.7 Normalized classifier output distributions of the Monte Carlo BDT . . . 61
7.8 ROC curves evaluating the performance of the Monte Carlo BDT classifiers on different backgrounds . . . 62
7.9 Normalized classifier output distributions of the QCD multijet BDT . . . 63
7.10 ROC curves evaluating the performance of the QCD multijet BDT classifiers on different backgrounds . . . 63
7.11 Fit regions of the $ZZ \rightarrow 4\tau$ cross section measurement . . . 66
7.12 Background composition and 1-bin data-to-model agreement of the different fit regions 67
7.13 Distribution of the $\log\left(\prod \tau^{\text{prob}}_{\text{i=1-3,4}}\right)$ observable for the constructed control regions . . 68
7.14 Qualitative pre-fit background composition in the CR_{Top} region . . . 69
7.15 Distribution of the $\log\left(\prod \tau^{\text{prob}}_{\text{i=1-3,4}}\right)$ observable for the constructed signal regions . . . 70
7.16 Distribution of the $\log\left(\prod \tau^{\text{prob}}_{\text{i=1-3,4}}\right)$ observable for the constructed validation region . 71

8.1 Distribution of the *tau_ID_eff_SYST* systematic uncertainty in the $\text{CR}_{\text{MCHighFLow}}$ region 75
8.2 Distribution of the *NP_MC* and *NP_Fakes* nuisance parameters in the $\text{CR}_{\text{MCLowFLow}}$ region . . . 76
8.3 Result of the pruning applied to all defined nuisance parameters . . . 77
8.4 Results for the POI, all nuisance parameters and the correlations between them for the fit on Asimov data . . . 78
8.5 Comparison of pre-fit and post-fit data-to-model agreement in the different fit regions for the fit on Asimov data . . . 79
8.6 Results for the POI, all nuisance parameters and the correlations between them for the fit on ATLAS 2017 data with an integrated luminosity of 46.9 fb^{-1} . . . 80
8.7 Best fit values for the gammas obtained from the fits on Asimov and ATLAS 2017 data 82

8.8 Post-fit data-to-model agreement in the different fit regions for the fit on ATLAS 2017 data . . . 83
8.9 Impact of the different nuisance parameters on the signal strength $\mu(ZZ \to 4\tau)$ . . . 84

B.1 Further input variables for the Monte Carlo BDT training (1/7) . . . 101
B.2 Further input variables for the Monte Carlo BDT training (2/7) . . . 102
B.3 Further input variables for the Monte Carlo BDT training (3/7) . . . 102
B.4 Further input variables for the Monte Carlo BDT training (4/7) . . . 103
B.5 Further input variables for the Monte Carlo BDT training (5/7) . . . 103
B.6 Further input variables for the Monte Carlo BDT training (6/7) . . . 104
B.7 Further input variables for the Monte Carlo BDT training (7/7) . . . 104
B.8 Further input variables for the Multijet BDT training (1/5) . . . 105
B.9 Further input variables for the Multijet BDT training (2/5) . . . 105
B.10 Further input variables for the Multijet BDT training (3/5) . . . 106
B.11 Further input variables for the Multijet BDT training (4/5) . . . 106
B.12 Further input variables for the Multijet BDT training (5/5) . . . 107

D.1 Overtraining results of the *BDT (MC)* classifier for different hyperparameter setups . 114
D.2 Overtraining results of the *BDTG (MC)* classifier for different hyperparameter setups 115
D.3 Overtraining results of the *BDT (Fakes)* classifier for different hyperparameter setups 116
D.4 Overtraining results of the *BDTG (Fakes)* classifier for different hyperparameter setups 117

List of Tables

2.1 Exchange particles in the Standard Model 4
2.2 Lorentz-invariant currents and their properties 6
2.3 Branching ratios of the Z boson into its different decay channels 9
2.4 Branching ratios of the Standard Model Higgs boson into its dominant decay channels 14

5.1 Identification efficiencies of $\tau_{\text{had-vis}}$ candidates and corresponding background rejection for the different defined tau working points 37

6.1 The two different DAOD types and their cut requirements used in this analysis 45
6.2 Preselection cuts placed on top of the DAOD cuts 45
6.3 Tau charge ratios for the signal processes, different backgrounds and recorded data . 47
6.4 Event yields of the different background processes resulting from the advanced data driven fake estimation on preselection level 50

7.1 Event yields of the $ZZ \rightarrow 4\tau$ signal, the background processes and recorded data after the cumulative application of a b-veto and different $\log\left(\prod \tau^{\text{prob}}_{\text{i=1-3,4}}\right)$ cuts 53
7.2 General training parameters of the boosted decision trees used for discrimination of the $ZZ \rightarrow 4\tau$ signal against its background 55
7.3 Number of training and test events of the boosted decision trees used for discrimination of the $ZZ \rightarrow 4\tau$ signal against its background 56
7.4 Classifier-specific top-10 ranked input variables for the training of the Monte Carlo BDT 58
7.5 Classifier-specific top-10 ranked input variables for the training of the QCD multijet BDT 58
7.6 Signal efficiency comparison between training and test sample at fixed background efficiencies for the different BDT classifiers 60
7.7 AUC values of the *BDT (MC)* and *BDTG (MC)* classifiers for different hyperparameter setups 62
7.8 AUC values of the *BDT (Fakes)* and *BDTG (Fakes)* classifiers for different hyperparameter setups 64
7.9 AUC values of the trained BDT classifiers evaluated on the different backgrounds . . 64
7.10 Labeling and selection criteria of the fit regions for the $ZZ \rightarrow 4\tau$ cross section measurement 66

8.1 Expected upper limit on $\mu(ZZ \rightarrow 4\tau)$ using Asimov data 79
8.2 Resulting post-fit values of the POI and the nuisance parameters using ATLAS 2017 data 81
8.3 Expected upper limit on $\mu(ZZ \rightarrow 4\tau)$ using ATLAS 2017 data 85

A.1 MC samples grouped into the Higgs background process as well as the $hh \rightarrow 4\tau$ signal sample . . . 97
A.2 MC samples grouped into the $t\bar{t}$ background process . . . 97
A.3 MC samples grouped into the Top background process . . . 98
A.4 MC samples grouped into the $Z \rightarrow \tau\tau$ background process . . . 98
A.5 MC samples grouped into the V + Jets background process . . . 99
A.6 MC samples grouped into the Multiboson background process . . . 100

C.1 Full classifier-specific rankings of all input variables for the training of the Monte Carlo BDT . . . 110
C.2 Full classifier-specific rankings of all input variables for the training of the QCD Multijet BDT . . . 111

E.1 Pre-Fit event yields in the $\mathrm{CR}_{\mathrm{MCLowFLow}}$ region . . . 119
E.2 Pre-Fit event yields in the $\mathrm{CR}_{\mathrm{MCHighFLow}}$ region . . . 119
E.3 Pre-Fit event yields in the $\mathrm{CR}_{\mathrm{Top}}$ region . . . 120
E.4 Pre-Fit event yields in the $\mathrm{SR}_{\mathrm{MCHighFMed}}$ region . . . 120
E.5 Pre-Fit event yields in the $\mathrm{SR}_{\mathrm{MCHighFHigh}}$ region . . . 121
E.6 Pre-Fit event yields in the $\mathrm{VR}_{\mathrm{MCLowFMed+}}$ region . . . 121

F.1 Post-Fit event yields in the $\mathrm{CR}_{\mathrm{MCLowFLow}}$ region . . . 123
F.2 Post-Fit event yields in the $\mathrm{CR}_{\mathrm{MCHighFLow}}$ region . . . 123
F.3 Post-Fit event yields in the $\mathrm{CR}_{\mathrm{Top}}$ region . . . 124
F.4 Post-Fit event yields in the $\mathrm{SR}_{\mathrm{MCHighFMed}}$ region . . . 124
F.5 Post-Fit event yields in the $\mathrm{SR}_{\mathrm{MCHighFHigh}}$ region . . . 124
F.6 Post-Fit event yields in the $\mathrm{VR}_{\mathrm{MCLowFMed+}}$ region . . . 124