Search for the Lepton-Flavor-Violating Higgs Boson Decays $H \to \tau\mu$ and $H \to \tau e$ in Di-Lepton Final States with the ATLAS Experiment

Masterarbeit von
Ulrich Baumann

betreut von
Prof. Dr. Markus Schumacher

Physikalisches Institut
Universität Freiburg
Hermann-Herder-Str. 3, 79104 Freiburg, Germany

4. November 2015

Abstract

A search for the lepton-flavor-violating decays $H \to \tau\mu$ and $H \to \tau e$ of the recently discovered Higgs boson with a mass of 125.09 GeV is performed. The presented analysis uses the full data set recorded by the ATLAS experiment at the LHC in 2012 at a center-of-mass energy of $\sqrt{s} = 8$ TeV corresponding to an integrated luminosity of $20.3\,\text{fb}^{-1}$. It focuses on the Higgs boson production mode of the vector-boson-fusion and considers hence only events in a topology enriched in this production mode. All fully leptonic decays of the τ-lepton ($\tau \to l\nu\bar{\nu}$ with $l = e, \mu$) in this vector-boson-fusion topology are analyzed. Events are split into four final states (ee, $\mu\mu$, $e\mu$ and μe) to increase sensitivity and to isolate both considered lepton-flavor-violating decays of the Higgs boson from each other. The analysis follows a multivariate approach using the score of Boosted Decision Trees as the final discriminant in each of the four final states. No hints for lepton-flavor-violating Higgs boson decays are observed. Model independent upper limits of 1.5 % (1.4 %) and 1.6 % (1.3 %) are observed (expected) for the branching ratio of the decays $H \to \tau\mu$ and $H \to \tau e$, respectively, at a 95 % confidence level.

Zusammenfassung

Eine Suche nach den leptonflavorverletzenden Zerfallsprozessen $H \to \tau\mu$ und $H \to \tau e$ des kürzlich entdeckten Higgs-Boson mit einer Masse von 125,09 GeV wird mit dem kompletten im Jahr 2012 aufgezeichneten Datensatz des ALTAS Experimentes am LHC durchgeführt. Dieser umfasst Proton-Proton-Kollisionen bei einer integrierten Luminosität von $20{,}3\,\text{fb}^{-1}$ und einer Schwerpunktsenergie von $\sqrt{s} = 8$ TeV. Die vorgestellte Analyse konzentriert sich auf den Produktionsmechanismus der Vektor-Boson-Fusion und berücksichtigt ausschließlich Ereignisse in einer Topologie, die in Vektor-Boson-Fusion produzierten Ereignissen angereichert ist. Alle vollständig leptonischen Zerfallsprozesse des τ Leptons ($\tau \to l\nu\bar{\nu}$ mit $l = e, \mu$) innerhalb dieser Topologie werden analysiert. Zur Erhöhung der Sensitivität und zur Trennung der berücksichtigten Zerfallsprozesse werden die Ereignisse in vier Endzustände (ee, $\mu\mu$, $e\mu$ und μe) unterteilt. Für jeden Endzustand wird mittels *Boosted Decision Trees* eine multivariate Analyse durchgeführt und die Verteilung deren *Score* als finale Diskriminante genutzt. Es werden keine Hinweise auf leptonflavorverletzende Zerfälle des Higgs-Bosons gefunden. Modellunabhängige obere Ausschlussgrenzen für das Verzweigungsverhältnis der betrachteten Zerfälle werden auf 95 % Konfidenzniveau abgeleitet. Diese beobachteten (erwarteten) Ausschlussgrenzen liegen bei 1,5 % (1,4 %) für den $H \to \tau\mu$ Zerfall und bei 1,6 % (1,3 %) für den $H \to \tau e$ Zerfall.

Erklärung

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

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

Contents

Chapter 1

Introduction

The Brout-Englert-Higgs mechanism [1–6], denoted from here on simply as the Higgs mechanism, was introduced more than 50 years ago. It allows to create mass terms for all massive particles without violating the gauge symmetry of the Standard Model (SM) [7–9]. The presence of the Higgs mechanism manifests itself in the existence of a scalar particle, the so called Higgs boson. While the SM does not predict the Higgs boson mass, it allows to infer all of its properties for a given value of its mass allowing purposeful searches for the Higgs boson. Although the experimental community put great effort into the search for this particle during the past 50 years by building and operating constantly larger particle accelerators pushing the frontier of the achievable energy-scale steadily higher, it was not discovered until 2012. In the meantime, all other particles predicted by the SM were discovered and the SM was found to describe physics at collider experiments with a remarkable precision, making the Higgs boson the last missing piece in the experimental confirmation of the SM.

In 2012, a candidate for the Higgs boson of the SM was discovered by the ATLAS [10] and CMS [11] collaborations at the LHC at CERN in the mass vicinity of 125 GeV. All measured properties of this candidate are so far in good agreement with the predictions of the SM, but further tests have to be carried out to identify the candidate as the Higgs Boson of the SM. The search for lepton flavor violation (LFV[1]) in the Higgs sector forms a intriguing class of such tests, due to multiple reasons.

First of all, the SM does not allow any LFV processes due to the properties of its lepton sector meaning that any observation of a LFV process calls for physics beyond the SM. It is nevertheless known that lepton flavor is not an exact symmetry of nature, as it is violated in the neutrino sector by the oscillation of the neutrinos' flavor during a free propagation in space [12–15] called neutrino oscillation [16–18]. Furthermore, many extensions of the SM [19–32] predict LFV in the Higgs sector. Indirect limits on the LFV Higgs boson decays, derived in references [33] and [34], allow the branching ratios of the decays $H \to \tau\mu$ and $H \to \tau e$ to be individually as large as $\mathcal{O}(10\,\%)$, while the branching ratio of the decay $H \to \mu e$ is in the absence of cancellations constrained to be smaller than $\mathcal{O}(10^{-8})$. This yields that there is not much hope for observing the decay $H \to \mu e$ with the ATLAS experiment at the LHC, while the decays $H \to \tau\mu$ and $H \to \tau e$ can be realized by nature at an observable rate. The CMS collaboration provided an additional motivation for a search of these LFV decays with the ATLAS experiment by observing a slight excess of events in the their search for the LFV Higgs decay $H \to \tau\mu$ [35] corresponding to a observed significance of $2.4\,\sigma$. The search for LFV Higgs decays is therefore a well motivated search for physics beyond the

[1] The abbreviation LFV will be used for both, *lepton flavor violation* and *lepton-flavor-violating*, trusting in the reader's skill to discriminate both options from context.

SM and tests furthermore the nature of the Higgs boson candidate.
So far, the ATLAS collaboration has only published a search for the LFV decay $H \to \tau\mu$ [36] considering only hadronic decays of the τ lepton. The search for the LFV decays $H \to \tau\mu$ and $H \to \tau e$ presented in this thesis considers, in contrast to this, the leptonic decay modes of the τ lepton. All fully leptonic final states are considered and the presented analysis exploits furthermore the distinct signature of the vector-boson-fusion (VBF) production mode of the Higgs boson by restricting the analysis to the VBF topology. It uses the full data set recorded by the ATLAS experiment at the LHC in 2012 at a center-of-mass energy of $\sqrt{s} = 8\,\text{TeV}$ corresponding to a integrated luminosity of $20.3\,\text{fb}^{-1}$. Events within the VBF topology are divided into four different final states, ee, $\mu\mu$, $e\mu$ and μe, to exploit the different signal and background compositions in each of those final states. The analysis is based heavily on the SM $H \to \tau\tau$ analysis [37] from the ATLAS collaboration and uses the same background estimates and object definitions. It uses however a different reconstruction of the Higgs boson mass and introduces the concept of lepton labeling to utilize the distinct kinematic of the LFV Higgs boson decays. The description of the performed analysis given in this thesis focuses on the parts of the analysis that differ from the SM $H \to \tau\tau$ analysis to emphasize the parts tailored to the LFV decays of the Higgs boson.
This thesis starts with a discussion of the theoretical background in chapter 2. This chapter gives an introduction to the SM focused on the absence of LFV in the SM, followed by a description of non SM LFV Higgs boson decays and a brief introduction to the phenomenology of the Higgs boson. A description of the ATLAS experiment is provided in chapter 3, which is followed by a discussion of the considered signal and background processes in chapter 4. The subsequent chapter 5 describes the reconstruction of physical objects, the used triggers and additionally the preselection of events defining the VBF topology. The concept of lepton labeling is introduced in chapter 6. This is followed by chapter 7 presenting the final event selection in all final states and the derivation of the distributions of the BDT score that are used as the final discriminant in the statistical analysis described in chapter 9. The intermediate chapter 8 describes briefly the used systematic uncertainties and how they are treated in this analysis. Eventually, conclusions are drawn in chapter 10.

Chapter 2

Theory

This chapter gives an overview of the main theoretical concepts underlying the analysis presented in this thesis. Starting with an introduction to the Standard Model (SM) of particle physics focused on the Higgs sector and the absence of LFV. The SMs predictions are used in this analysis for the modeling of all Monte Carlo based background simulations and, to some extent, in the simulation of the LFV Higgs boson signal samples. The section on the SM is followed by a section dedicated to LFV in the Higgs sector, describing a simple way to extend the Lagrangian of the SM such that it includes LFV decays of the Higgs boson. Furthermore, it contains a discussion of the indirect limits on LFV decays of the Higgs boson. Finally, the chapter is completed by a section about the Higgs boson itself. It describes the main properties of the SM Higgs boson and of LFV Higgs boson decays and gives a summary of the properties of the Higgs boson candidate discovered at a mass of approximately 125 GeV.

2.1 The Standard Model of Particle Physics

The Standard Model (SM) of particle physics describes all currently scientifically proven elementary particles, their relevant[1] interactions and properties. Gravitation is the only known fundamental force that is not included in the SM, as it is very weak compared to all other known fundamental forces and therefore negligible at the energy scale reached by today's experiments.
The SM contains two fundamentally different classes of particles, the fermions and the bosons. The criteria that sorts particles in these two groups is their intrinsic spin. Fermions are particles with a half integer spin quantum number, while bosons have an integer spin. All known elementary fermions have spin 1/2 and are divided in two subclasses depending on their color charge. Quarks are color charged elementary fermions and are therefore affected by the strong interaction, while leptons are neutral with respect to the color charge. Both, leptons and quarks, come in two different types. There are electrically charged leptons called electron, muon, tau (e, μ, τ), that carry the negative of the elementary charge e and their electrically neutral partners called neutrinos (ν_e, ν_μ, ν_τ) that are assumed to be massless within the SM. The quark sector is divided in so called up- and down-type quarks. There are three up-type quarks *up*, *charm* and *top* (u, c, t) that carry the electric charge $+2/3\,e$ and the three down-type quarks called *down*, *strange* and *bottom* (d, s, b) carrying the electric charge $-1/3\,e$. The SM contains each quark three times, once in each of the

[1] Interactions that are strong enough to leave a visible footprint at the energy scale of current particle accelerators.

three different possible color charges *red*, *blue* and *green*. For each of the so far described fermions exists additionally a corresponding anti-fermion having the same mass, spin and absolute value of electric charge with a sign opposite to the corresponding fermion.
All leptons and quarks of a certain type (charged leptons, neutrinos, up-type and down-type quarks) have almost the same properties. This manifests itself in the terminology of fermion generations. There are exactly three generations of each fermion type. Fermions of the same type but a different generation differ only in their mass and flavor. Each fermion class (leptons and quarks) has one independent flavor quantum number for each generation, e and ν_e for example have the same flavor quantum number. Flavor quantum numbers are additive quantum numbers and the flavor quantum number of an anti-particle is minus the flavor quantum number of the corresponding particle. The SM does not allow a violation of the leptons' flavor quantum numbers, or in short lepton flavor violation, while the conservation of quark flavor is violated by the charged weak interaction.
Matter, as we know it, is composed completely out of the three lightest fermions of each charged fermion type which are the electron, the *up* and *down* quarks. The two quarks form the protons (uud) and the neutrons (udd) building the nuclei that are completed by the electrons to form atoms. All other charged elementary particles are unstable and will decay.
Bosons are the force carriers of the SM mediating the elementary forces between the elementary particles. All elementary bosons of the SM have spin 1 except the Higgs boson which is with its spin 0 a unique elementary particle in the SM. The photon γ is the mediator of the electromagnetic force and couples to electric charge. It is neutral with respect to the color, the weak and electric charge and does not couple to itself due to the latter. The strong interaction is mediated by eight color charged but electric neutral gluons g that couple to themselves. Both, the photon and the gluon are massless in contrast to the massive Z, W^+ and W^- boson that are the mediators of the weak interaction. The bosonic content of the SM is completed by Higgs boson. It is massive scalar (spin 0) that couples to all massive particles of the SM. A summary of all particles of the SM and their main properties is given in Figure 2.1.
All charged fermions and spin 1 bosons are experimentally well established and are described precisely by the SM[2]. A candidate for the SM Higgs boson has been discovered by ATLAS [10] and CMS [11] at the LHC at CERN in 2012 in the mass vicinity of 125 GeV. All measured properties are within their uncertainties in good agreement with the SM predictions, but further, more accurate measurements will have to be carried out to gain enough confidence in stating that this candidate is in fact the Higgs boson of the SM.
While the SM describes precisely the physics of the elementary particles at colliders, it is known to have some limitations. E.g. it does not include gravitation, offers no candidate for dark matter, does not explain why there are exactly three generations of fermions and why the fermion masses span over so many orders of magnitude.
From a theoretical and therefore mathematical point of view, the SM is formulated as local gauge invariant quantum field theory (QFT) with the gauge structure $SU(3)_C \times SU(2)_{I_W} \times U(1)_{Y_W}$ composed of the $SU(3)_C$ group of color, the $SU(2)_{I_W}$ group of weak isospin I_W and the $U(1)_{Y_W}$ group of weak hypercharge Y_W. Quantum field theories describe particles as quantized fields, while obeying the rules of special relativity and quantum mechanics. Naive mass-terms for the particle content of the SM would break local gauge invariance, make the theory unrenormalizable and therefore cannot be included in the SM. The Higgs mechanism provides an elegant way to create mass terms for the massive particles of the SM without violating gauge invariance and destroying the renormalizability of the theory. It introduces the scalar Higgs field that has a non-vanishing

[2]Neutrinos are purposely excluded from this statement as neutrino oscillations are not described by the SM but experimentally well established.

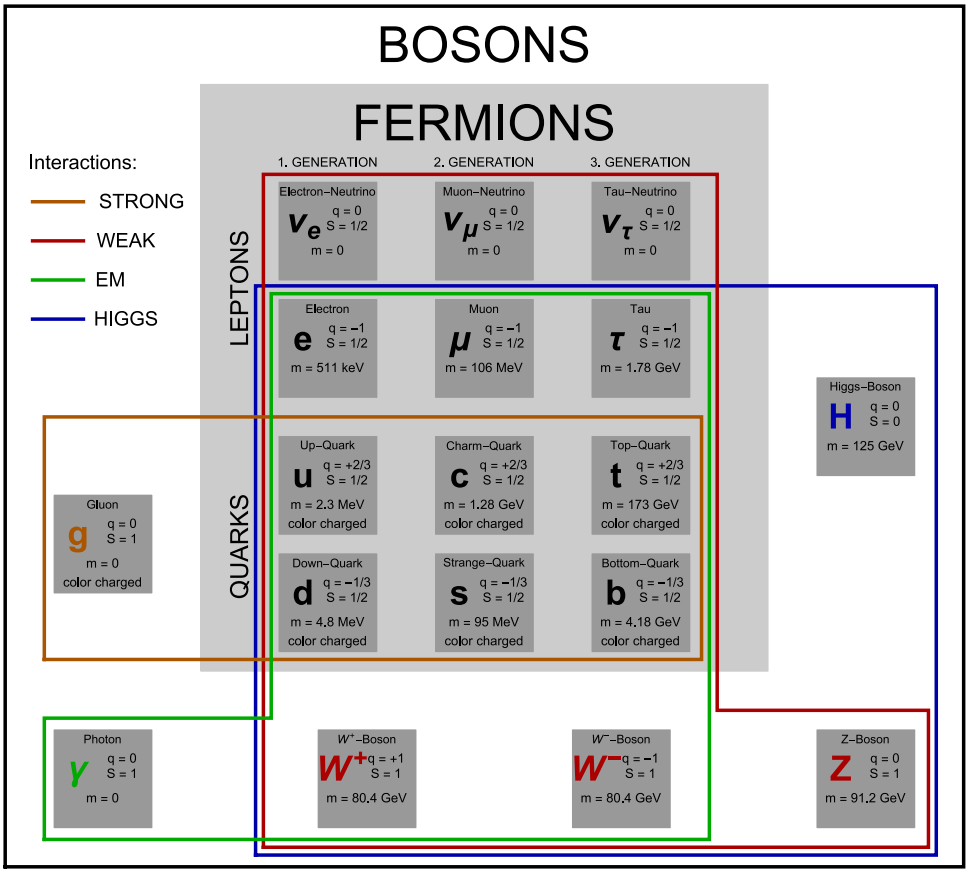


Figure 2.1: Summary of the SM particle content and its fundamental interactions represented by the different colors. Bosons with the same color as the box are the force carriers of the corresponding interaction, while all other particles inside each colored box are affected by the interaction represented by the color. It lists the electric charge q, the spin S and the mass m of the particles. The mass values are taken from reference [38].

vacuum expectation value and breaks therefore spontaneously the $SU(2)_{I_W} \times U(1)_{Y_W}$ symmetry. Due to this, interaction terms arise that are bilinear in the massive vector boson fields and the vacuum expectation value of the Higgs fields that can be interpreted as mass terms. Fermion masses are obtained separately by the so-called Yukawa interactions.
This section will provide a phenomenological description of the SM focusing on the Higgs sector and the absence of LFV processes. It is based heavily on the elegant summary of the SM given in reference [39] and follows mostly the same notation. Furthermore, a certain familiarity of the reader with the basics of QFT, the usual notations and tools in this field is assumed from here on, as an introduction to QFT is beyond the scope of this thesis.
The Lagrangian of the Standard Model $\mathcal{L}_{\text{SM}}$ given by

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{YM}} + \mathcal{L}_{\text{F}} + \mathcal{L}_{\text{H}} + \mathcal{L}_{\text{Yuk}}. \tag{2.1}$$

$\mathcal{L}_{\text{SM}}$ will be discussed in detail in the following sections, starting with $\mathcal{L}_{\text{YM}}$ and $\mathcal{L}_{\text{F}}$.

2.1.1 $\mathcal{L}_{\mathbf{YM}}$ and $\mathcal{L}_{\mathbf{F}}$

$\mathcal{L}_{\mathrm{YM}}$ and $\mathcal{L}_{\mathrm{F}}$ describe both the fermions' and gauge bosons' dynamics and their interactions with each other in an unbroken gauge symmetry. The Lagrangian of the Yang-Mills part is given by

$$\mathcal{L}_{\mathrm{YM}} = -\frac{1}{4}W^i_{\mu\nu}W^{i,\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}G^a_{\mu\nu}G^{a,\mu\nu}, \tag{2.2}$$

using the following definitions of the field strength tensors

$$W^i_{\mu\nu} = \partial_\mu W^i_\nu - \partial_\nu W^i_\mu - g\epsilon^{ijk}W^j_\mu W^k_\nu, \qquad i,j,k = 1,2,3, \tag{2.3}$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu, \tag{2.4}$$

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - g_s f^{abc}G^b_\mu G^c_\nu, \qquad a,b,c = 1,2,...,8. \tag{2.5}$$

Here, the $W^i_{\mu\nu}$ denote the field strength tensors of the gauge fields W^i_μ for the $SU(2)_{I_W}$ group of weak isospin I^i_W. $B_{\mu\nu}$ is the field strength tensor of the gauge field B_μ for the $U(1)_{Y_W}$ of weak hypercharge Y_W and finally $G^a_{\mu\nu}$ are the field strength tensors for the $SU(3)_C$ of color. The gauge couplings of these groups are given by g, g' and g_s respectively. Additionally, there are the structure constants ϵ^{ijk} and f^{abc} of the non-abelian groups $SU(2)_{I_W}$ and $SU(3)_C$ defined by the commutation relations of the groups' generators.
Equation 2.2 describes the dynamics of the gauge bosons and their self interaction, while

$$\mathcal{L}_{\mathrm{F}} = i\bar{\Psi}_L \not{D}\Psi_L + i\bar{\psi}_{l_R}\not{D}\psi_{l_R} + i\bar{\Psi}_Q\not{D}\Psi_Q + i\bar{\psi}_{u_R}\not{D}\psi_{u_R} + i\bar{\psi}_{d_R}\not{D}\psi_{d_R} \tag{2.6}$$

describes the fermions' dynamics and the interaction between fermions and gauge bosons, which is hidden in the covariant derivative D_μ given by

$$D_\mu = \partial_\mu + igI^i_w W^i_\mu + ig'\frac{Y_W}{2}B_\mu + ig_s T^a_C G^a_\mu. \tag{2.7}$$

$\not{D}_\mu$ is the Feynman slashed of D_μ defined by $\not{D} = D_\mu\gamma^\mu$. I^i_W, Y_W, and T^a_C are the generators of the gauge groups in the representation of the fermions they act on. $I^i_W = \frac{\sigma^i}{2}$ (σ^i = Pauli matrices) for left-handed $SU(2)_{I_W}$ doublets and $I^i_W = 0$ for right-handed singlets. A direct consequence of this is that the $W^\pm$ boson of the SM couples only to left-handed fermions which makes the charged weak current maximal parity violating. The weak hypercharge Y_W is related to the relative electric charge Q and the third component of the weak isospin I^i_W by the Gell-Mann-Nishijima relation $Q = I^3_W + \frac{Y_W}{2}$. $T^a_C = \frac{\lambda^a}{2}$ (λ^a = Gell-Mann matrices) for $SU(3)_C$ quark triplets and $T^a_C = 0$ for leptons.
Equation 2.6 introduced the SM fermion content that will be discussed below. The SM treats left- and right-handed fermions differently. The left-handed fermions appear in $SU(2)_{I_W}$ doublets ($I_W = \pm 1/2$), while the SM contains their right-handed counterparts only in singlets ($I_W = 0$). There appear two doublets in Equation 2.6: the left-handed lepton doublets[3] $\Psi_L = (\psi_{\nu_{l_L}}, \psi_{l_L})^T$ consisting of the spinor ψ_{l_L} of the charged leptons $l = e, \mu, \tau$ and the spinor $\psi_{\nu_{l_L}}$ of the corresponding neutrinos $\nu_l = \nu_e, \nu_\mu, \nu_\tau$, and the left-handed quark doublet $\Psi_Q = (\psi_{u_L}, \psi_{d_L})^T$ containing the spinor ψ_{u_L} of the up-type quarks $u = u, c, t$ and the spinor ψ_{d_L} of the down-type quarks $d = d, s, b$. Furthermore, there are the singlets of the right-handed charged leptons ψ_{l_R}, of the right-handed

[3] please note that the subscript L is used twice in different context: Once in labeling Ψ_L as the doublet containing left-handed fermions (as Ψ_Q labels the left-handed quark doublet) and once as labeling the left-handedness of spinors like ψ_{l_L}. With this in hand, it should be easy to infer the proper meaning of the subscript L from context.

up-type quarks ψ_{u_R} and of the right-handed down-type quarks ψ_{d_R}. Note, there is no singlet of the right-handed neutrinos, as the SM does not contain right-handed neutrinos.
So far, the gauge fields were treated in their gauge basis and not as usual in experimental particle physics in the mass basis ($W_\mu^\pm$,Z_μ, A_μ) with their corresponding particles ($W^\pm$, Z, γ). Requiring that the coupling structure of the photon in the SM is the same as in quantum electrodynamics (parity blind and proportional to $Q\bar{\psi}\not{A}\psi$) allows to identify the photon field A_μ and the Z boson field Z_μ as a rotation of the W_μ^3 and B_μ fields given by

$$\begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} c_W & -s_W \\ s_W & c_W \end{pmatrix} \begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix}. \tag{2.8}$$

The angle of the rotation is the weak mixing angle θ_W. The abbreviations c_W and s_W are defined as

$$c_W = \cos\theta_W = \frac{g}{\sqrt{g^2+g'^2}} \qquad s_W = \sin\theta_W = \frac{g'}{\sqrt{g^2+g'^2}}. \tag{2.9}$$

The electric charge e is fixed by

$$e = \frac{gg'}{\sqrt{g^2+g'^2}}. \tag{2.10}$$

Furthermore, the $W_\mu^\pm$ fields can be identified by

$$W_\mu^\pm = \frac{1}{\sqrt{2}}\left(W_\mu^1 \mp iW_\mu^2\right). \tag{2.11}$$

This section introduced all particles of the SM except the scalar Higgs boson, but so far all particles remained massless, as naive mass terms such as $W_\mu^i W^{i,\mu}$ for gauge bosons and $\bar{\psi}_{f_L}\psi_{f_R} + \bar{\psi}_{f_R}\psi_{f_L}$ for fermions violate gauge invariance.

2.1.2 Higgs Part of the SM

This section introduces the Higgs sector of the SM that provides the possibility of creating mass terms for all massive particles without violating the gauge invariance of the theory. The Lagrangian of the Higgs sector is given by

$$\mathcal{L}_\mathrm{H} = (D_\mu\Phi)^\dagger (D^\mu\Phi) - V(\Phi), \tag{2.12}$$

introducing a complex scalar $SU(2)_{I_W}$ doublet $\Phi = (\phi^+, \phi^0)^T$ with four degrees of freedom. Φ is neutral with respect to the color charge and carries the weak hypercharge $Y_W = 1$. ϕ^+ carries additionally the charge $+e$, while ϕ^0 is neutral. D_μ is the covariant derivative introduced in equation 2.7. The potential V of the self interaction of Φ is given by

$$V(\Phi) = -\mu^2\left(\Phi^\dagger\Phi\right) + \frac{\lambda}{4}\left(\Phi^\dagger\Phi\right)^2, \tag{2.13}$$

where μ^2 and λ are two real free parameters. μ^2 is chosen positive so that Φ picks up a non-vanishing vacuum expectation value (vev) Φ_0, which is crucial to create mass terms for massive particles using the Higgs mechanism. Minimizing V with respect to $\left(\Phi^\dagger\Phi\right)$ yields

$$\Phi_0^\dagger\Phi_0 = \frac{v^2}{2}, \tag{2.14}$$

where

$$v = 2\sqrt{\frac{\mu^2}{\lambda}} \tag{2.15}$$

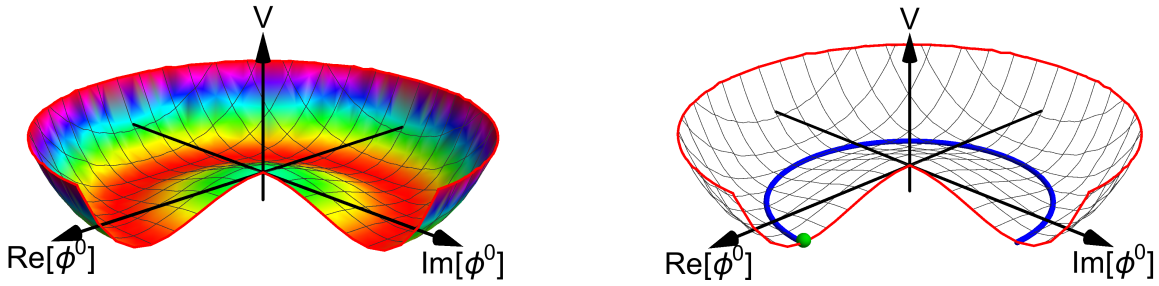
denotes the vev. As ϕ^+ carries the charge e, it is necessary to set $\phi_0^+ = 0$ to keep the vev and therefore the vacuum neutral. With this requirement Φ_0 is fixed up to an arbitrary phase. Following the usual choice and not adding an additional phase yields $\Phi_0 = \left(0, \frac{v}{\sqrt{2}}\right)^T$. This choice of the vev breaks the initial $SU(2)_{I_W} \times U(1)_{Y_W}$ symmetry spontaneously down to the remaining $U(1)_{EM}$ symmetry. Rewriting Φ in terms of v, the real physical Higgs field H, and the unphysical would-be Goldstone bosons ϕ^+ (complex scalar) and χ (real scalar) yields

$$\Phi = \begin{pmatrix} \phi^+ \\ \frac{v+H+i\chi}{\sqrt{2}} \end{pmatrix}. \tag{2.16}$$

It is always possible to find a gauge that removes the unphysical fields ϕ^+ and χ from Φ, known as the *unitary gauge*. In this gauge Φ reduces to

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H \end{pmatrix}. \tag{2.17}$$

Figure 2.2 shows the potential V projected onto the ϕ^0 component. The blue line in Figure 2.2 (b) shows the continuum of vacuum expectation values and the green dot represents selected vev.



(a) (b)

Figure 2.2: Two plots of the Higgs potential V projected to the field component ϕ^0 of Φ. The blue circle in plot (b) shows the continuum of the vacuum expectation values, while the green dot represents the selected vev.

The interaction between Φ and the gauge boson fields of the SM is hidden in the covariant derivative in equation 2.12. Expanding the covariant derivative given in equation 2.7 with $I^i_{W,\Phi} = \sigma^i$, $Y_{W,\Phi} = 1$, $T^a_{C,\Phi} = 0$ and replacing W^i_μ and B_μ with the corresponding fields in the mass basis $W^\pm_\mu$, A_μ and Z_μ simplifies the covariant derivative to

$$D_\mu = \partial_\mu + \frac{i}{2} \begin{pmatrix} \frac{g^2 - g'^2}{\sqrt{g^2+g'^2}} Z_\mu + 2eA_\mu & \sqrt{2} g W^+_\mu \\ \sqrt{2} g W^-_\mu & -\frac{g}{c_W} Z_\mu \end{pmatrix}. \tag{2.18}$$

Using this into calculate $\mathcal{L}_H$ and the unitary gauge yields

$$\mathcal{L}^U_H = \tfrac{1}{2}\partial_\mu H \partial^\mu H + \tfrac{g^2}{4}(v+H)^2 W^+_\mu W^{-\mu} + \tfrac{g^2}{8c_W^2}(v+H)^2 Z_\mu Z^\mu + \tfrac{\mu^2}{2}(v+H)^2 - \tfrac{\lambda}{16}(v+H)^4. \tag{2.19}$$

This contains terms bilinear in the fields $W^{\pm}$, Z^{μ} and H, which can be identified as mass terms giving the corresponding bosons ($W^{\pm}$, Z, and H) their mass M. These mass terms are given by

$$M_W = \tfrac{gv}{2}, \qquad M_Z = \tfrac{gv}{2c_W} = \tfrac{M_W}{c_W}, \qquad M_H = v\sqrt{\tfrac{\lambda}{2}} = \sqrt{2\mu^2}. \tag{2.20}$$

Expanding equation 2.19 and replacing the parameters μ^2, v, λ with the bosons' masses yields

$$\begin{aligned}\mathcal{L}_H^U =& \tfrac{1}{2}\partial_\mu H \partial^\mu H - \tfrac{1}{2}M_H^2 H^2 - \tfrac{gM_H^2}{4M_W}H^3 - \tfrac{g^2 M_H^2}{32M_W^2}H^4 \\ &+ M_W^2 W_\mu^+ W^{-,\mu} + \tfrac{1}{2}M_Z^2 Z_\mu Z^\mu \\ &+ gM_W H W_\mu^+ W^{-,\mu} + \tfrac{g^2}{4}H^2 W_\mu^+ W^{-,\mu} + \tfrac{gM_Z}{2c_W}H Z_\mu Z^\mu + \tfrac{g^2}{4c_W^2}H^2 Z_\mu Z^\mu.\end{aligned} \tag{2.21}$$

The first line of equation 2.21 describes the kinematics and self-interaction of the real, scalar and now massive Higgs field that inherited one of the four degrees of freedom of the initial Higgs doublet Φ. The residual three degrees of freedom are transfered to the fields W_μ^+, W_μ^- and Z_μ, as they became massive due the mass term given in the second line of the equation. Each of them picks up one additional degree of freedom, as massive fields with spin one have three possible spin states, while massless fields with spin one have only two spin states. The last line of equation 2.21 describes the couplings between H and the massive gauge bosons.
The introduction of $\mathcal{L}_H$ allows to create mass terms for the $W_\mu^{\pm}$ and Z_μ fields, while keeping the photon field A_μ massless, which matches the observation of massive $W^{\pm}$, Z boson and a massless photon γ in nature, without breaking the gauge invariance of the theory. This achievement comes with the price of having to add the Higgs boson H to the particle content of the SM interacting with all massive gauge bosons.
$\mathcal{L}_H$ introduced two free parameters μ^2 and λ. They are both fixed with the input of the Higgs boson mass and the vev $v = 247\,\text{GeV}$ derived from the Fermi constant.

2.1.3 Yukawa Couplings

The SM includes interactions between fermions and the Higgs sector realized by so called Yukawa [40, 41] interactions. From these interactions arise the fermions' mass terms and couplings between fermions and the Higgs boson. This is achieved by coupling a left-handed fermion doublet Ψ with a right-handed fermion singlet ψ to the Higgs doublet Φ. The most general Lagrangian for these couplings is given by

$$\mathcal{L}_{\text{Yuk}} = -\underbrace{\bar{\Psi}_L G_l \psi_{l_R}\Phi}_{A} - \underbrace{\bar{\Psi}_Q G_d \psi_{d_R}\Phi}_{B} - \underbrace{\bar{\Psi}_Q G_u \psi_{u_R}\tilde{\Phi}}_{C} + \text{h.c.}, \tag{2.22}$$

where G_f, with $(f = l, u, d)$, denotes an arbitrary complex 3x3 matrix in flavor space containing the coupling constants G_{ij} between the fermion flavors of the lepton generation i and j and $\tilde{\Phi} = i\sigma^2\Phi^*$ denotes the charge-conjugate Higgs doublet.
Part A of equation 2.22 describes the Yukawa interaction for charged leptons, part B for *down-type* quarks and part C for *up-type quarks*. The SM Yukawa couplings treat leptons and *down-type* quarks analogously, as can be seen easily by comparing Part A and B, but there is no corresponding term to part C for leptons. This is due to the fact that the SM does not contain right-handed neutrinos (and left-handed anti-neutrinos) which are essential to form the corresponding term for leptons. This difference between quarks and leptons leads to a different behavior of quarks and leptons in

terms of flavor violation, which will be shown in section 2.1.4. Expanding equation 2.22 in the unitary gauge yields

$$\mathcal{L}_{\text{Yuk,U-gauge}} = \frac{1}{\sqrt{2}}(v+H)\left(-\bar{\psi}_{l_L} G_l \psi_{l_R} - \bar{\psi}_{d_L} G_d \psi_{d_R} - \bar{\psi}_{u_L} G_u \psi_{u_R}\right) + \text{h.c.}. \tag{2.23}$$

The expanded version of equation 2.22 contains terms that are already quite similar to mass terms and Higgs-fermion interactions, but in general the arbitrary G-matrices mix different lepton (quark) flavors with each other. This mixture would lead to an oscillation of a lepton (quark) with flavor f_i of the i^{th} generation (i=1,2,3) into a lepton (quark) with flavor f_j ($i \neq j$) simply by free propagation of the lepton (quark) in time and space. It is not very convenient to keep those flavor mixing terms in the SM as they complicate the theory unnecessarily. Fortunately, it is possible to remove them by a simple redefinition of the fermions' fields shown in the next section.

Fermion Masses and Higgs Boson Couplings to Fermions

Changing the fermions' basis from their previously used *flavor basis* ($\psi_{\pi,1}$,$\psi_{\pi,2}$,$\psi_{\pi,3}$) into their *mass basis* ($\hat{\psi}_{\pi,1}$,$\hat{\psi}_{\pi,2}$,$\hat{\psi}_{\pi,3}$) for both, left-handed ($\pi = L$) and right-handed ($\pi = R$), fermions provides a convenient method to eliminate flavor mixing terms from equation 2.23. This change of basis can be expressed as

$$\hat{\psi}_{f_{\pi,i}} = U_{ij}^{f_\pi} \psi_{f_{\pi,j}}, \qquad f = l, u, d, \tag{2.24}$$

using the unitary matrices $U^{f,\pi}$ that diagonalize G_f by definition of the *mass basis*. Please note that here it is necessary to introduce two different *types* of matrices, namely U^{u_π} and U^{d_π}, for the quarks sector while there is only U^{l_π} for the lepton sector, as only the charged leptons remained in $\mathcal{L}_{\text{Yuk}}$, while both, up- and down-type of quarks, are still present. Although it is not necessary to introduce the corresponding U matrix for the neutrinos at this stage, it will be crucial later, once the change of the fermion basis is performed on the entire Lagrangian of the SM. So it is convenient to define the transformation matrix for the neutrinos at this stage. Fortunately, it is possible to use U^{l_π} for both the charged leptons and the neutrinos, while working in the approximation of massless and therefore mass degenerate neutrinos, as in the SM.
A notationally convenient choice for the parametrization using the actual fermions mass m_f of the mass eigenvalues is given by

$$U^{f_L} G_f \left(U^{f_R}\right)^\dagger = \frac{\sqrt{2}}{v} diag\{m_{f_1}, m_{f_2}, m_{f_3}\}. \tag{2.25}$$

By applying this notation and using $\bar{\psi}_{f_\pi} = \bar{\hat{\psi}}_{f_\pi} \left(U^{f_\pi}\right)^\dagger$, equation 2.23 reads

$$\mathcal{L}_{\text{Yuk,U-gauge}} = -\sum_f \sum_i m_{f_i} \left(1 + \frac{H}{v}\right) \left(\bar{\hat{\psi}}_{f_{i_L}} \hat{\psi}_{f_{i_R}} + \bar{\hat{\psi}}_{f_{i_R}} \hat{\psi}_{f_{i_L}}\right), \tag{2.26}$$

where the subscript i represents the fermion's generation and f the fermion's type (l,u,d). Equation 2.26 contains two types of terms. First of all, there are the fermions' mass terms providing each prior massless fermion with its correct mass m_f. Furthermore, there are terms describing a coupling between each fermion and the Higgs boson of the SM, that is proportional to $\frac{m_f}{v}$. This means that the Higgs boson couples only to massive fermions and favors heavy fermions in its decay. Additionally, there are no terms left in equation 2.26 that mix fermions of different flavor with each other. Thus, the Higgs boson of the SM can, at tree level, only decay into pairs (fermion and anti-fermion) of massive fermions with the same flavor, considering only fermionic decays. So,

the SM does not allow any lepton flavor violating decays of the Higgs boson and observing such a decay would call for physics beyond the SM. Obviously, it is not possible to change the basis of the fermions in the Yukawa sector of the SM without changing it in the whole SM. Therefore, it is necessary to switch from the fermions' flavor basis into their mass basis in the entire SM simultaneously. The consequences of this will be discussed in the next section.

2.1.4 Flavor Violation in the Quark Sector and Absence of Lepton-Flavor-Violating Processes in the Standard Model

Fermions appear twice in the SM, once in the Yukawa couplings $\mathcal{L}_{\text{Yuk}}$ and once in $\mathcal{L}_F$ given in equation 2.6. As the effects of the change of the fermions' basis on $\mathcal{L}_{\text{Yuk}}$ were already discussed in detail in the previous section there is with $\mathcal{L}_F$ only one place left to enforce this change of basis. Carrying out the change of basis in $\mathcal{L}_F$ yields

$$\begin{aligned}\mathcal{L}_{\text{F}} =& i\left(\bar{\hat{\psi}}_{\nu_{lL}} U^{l_L}, \bar{\hat{\psi}}_{l_L} U^{l_L}\right) \not{D} \begin{pmatrix} \left(U^{l_L}\right)^\dagger \hat{\psi}_{\nu_{lL}} \\ \left(U^{l_L}\right)^\dagger \hat{\psi}_{l_L} \end{pmatrix} + i\bar{\hat{\psi}}_{l_R} U^{l_R} \not{D} \left(U^{l_R}\right)^\dagger \hat{\psi}_{l_R} \\ &+ i\left(\bar{\hat{\psi}}_{u_L} U^{u_L}, \bar{\hat{\psi}}_{d_L} U^{d_L}\right) \not{D} \begin{pmatrix} \left(U^{u_L}\right)^\dagger \hat{\psi}_{u_L} \\ \left(U^{d_L}\right)^\dagger \hat{\psi}_{d_L} \end{pmatrix} \\ &+ i\bar{\hat{\psi}}_{u_R} U^{u_R} \not{D} \left(U^{u_R}\right)^\dagger \hat{\psi}_{u_R} + i\bar{\hat{\psi}}_{d_R} U^{d_R} \not{D} \left(U^{d_R}\right)^\dagger \hat{\psi}_{d_R}.\end{aligned} \tag{2.27}$$

Here, it is worth stressing once more that the neutrinos are transformed with the same matrix as the charged leptons, which is possible in the approximation of mass degenerate neutrinos as discussed in the previous section.
Equation 2.27 can be simplified easily by using the unitarity of the U^{f_π} matrices and the fact that $\not{D}$ and U^{f_π} commute as U^{f_π} is defined in the flavor space and $\not{D}$ is proportional to the identity in flavor space. With this in hand it is easy to see that all U^{f_π} and $(U^{f_\pi})^\dagger$ matrices cancel each other $(U^{f_\pi}(U^{f_\pi})^\dagger = \mathbb{1})$ for all terms containing the right-handed fermions' singlets and for the left-handed lepton doublet. Applying this to $\mathcal{L}_F$ yields

$$\begin{aligned}\mathcal{L}_{\text{F}} =& i\bar{\hat{\Psi}}_L \not{D} \hat{\Psi}_L + i\bar{\hat{\psi}}_{l_R} \not{D} \hat{\psi}_{l_R} + i\left(\bar{\hat{\psi}}_{u_L} U^{u_L}, \bar{\hat{\psi}}_{d_L} U^{d_L}\right) \not{D} \begin{pmatrix} \left(U^{u_L}\right)^\dagger \hat{\psi}_{u_L} \\ \left(U^{d_L}\right)^\dagger \hat{\psi}_{d_L} \end{pmatrix} \\ &+ i\bar{\hat{\psi}}_{u_R} \not{D} \hat{\psi}_{u_R} + i\bar{\hat{\psi}}_{d_R} \not{D} \hat{\psi}_{d_R},\end{aligned} \tag{2.28}$$

where the notation $\hat{\Psi}_L = \left(\hat{\psi}_{\nu_{lL}}, \hat{\psi}_{l_L}\right)$ was introduced. The term for the left-handed quarks was left untouched. This is due to the fact that the same procedure does not work for the left-handed quark doublet, as up- and down-type quarks transform with respect to two different matrices U^{u_π} and U^{d_π}. The off diagonal elements of $\not{D}$ mix $\hat{\psi}_{u_L}$ and $\hat{\psi}_{d_L}$ with each other causing bilinear terms in U^{u_L} and U^{d_L}. As these matrices are in general not identical and therefore cannot cancel each other, there will be a residual effect of the change of basis in the quark sector on the off diagonal elements of $\not{D}$. The diagonal elements will be left unmodified, as there appears no mixture of U^{u_L} and U^{d_L}. The off-diagonal elements in $\not{D}$ are the source of the charged weak current transmitted by the $W^\pm$ boson exchange. The interaction of the $W^\pm$ bosons with the left-handed quarks is given

by

$$\begin{aligned}\mathcal{L}_F|_{\text{Q-}W^\pm\text{-Int.}} &= -\tfrac{1}{2}\left(\bar{\hat{\psi}}_{u_L}U^{u_L}, \bar{\hat{\psi}}_{d_L}U^{d_L}\right)\gamma^\mu\begin{pmatrix}0 & \sqrt{2}gW_\mu^+\\ \sqrt{2}gW_\mu^- & 0\end{pmatrix}\begin{pmatrix}(U^{u_L})^\dagger\hat{\psi}_{u_L}\\ (U^{d_L})^\dagger\hat{\psi}_{d_L}\end{pmatrix}\\ &= -\frac{1}{\sqrt{2}}g\left(\bar{\hat{\psi}}_{u_L}V\gamma^\mu W_\mu^+\hat{\psi}_{d_L} + \bar{\hat{\psi}}_{d_L}V^\dagger\gamma^\mu W_\mu^-\hat{\psi}_{u_L}\right),\end{aligned} \tag{2.29}$$

where the Cabibbo-Kobayashi-Maskawa (CKM) [42, 43] matrix $V = U^{u_L}\left(U^{d_L}\right)^\dagger$ was introduced. The CKM matrix is the only left-over of the change into the fermions mass basis in the entire SM. There is no prediction of the elements of the CKM matrix by the SM which means that they have to be determined in experiment. In fact, it is found to be non-diagonal, such that it mixes quarks of different flavor in the charged weak interaction and therefore, there is no conservation of quark flavor in the SM.
The equivalent to the CKM matrix in the lepton sector of the SM is the identity matrix, as neutrinos and charged leptons transform in the SM under the same matrix U. So, there is no mixture of leptons in the weak interaction and therefore no source of lepton flavor violation due to the change of the fermions' basis. In summary, it was possible to remove the lepton flavor violating terms from the Yukawa interactions by transforming the fermions into their mass basis and this change of basis did not introduce any lepton flavor violating processes in the SM. So, lepton flavor is conserved in the SM and any observation of a violation of lepton flavor conservation calls for physics beyond the SM.

2.2 Lepton-Flavor-Violation in non Standard Model Higgs Boson Decays

The SM does not allow any LFV processes as shown in section 2.1.4. There is nevertheless a LFV process known to exist in nature, which is neutrino oscillation [16–18]. Neutrino oscillation describes the phenomenon that the flavor of a neutrino oscillates during a free propagation in space meaning that a neutrino created with the flavor of the i^{th} generation at the location A has in general a non vanishing probability to be measured as a neutrino of the flavor of the j^{th} generation at the location B ($i \neq j$ and $A \neq B$). This was experimentally observed e.g. by references [12–15]. Neutrino oscillation is only possible if the masses of the neutrinos are not degenerate meaning that at lest two of the three known neutrinos must be massive contradicting the assumptions of massless neutrinos of the SM.
Due to this, lepton flavor cannot be an exact symmetry of nature contradicting the SM prediction. Many extensions of the SM, like models containing more than one Higgs doublet [19–22] and many others [23–32], predict LFV processes in the Higgs sector. Specific models are not discussed in this thesis, as the analysis presented is model independent and a discussion of specific extensions is beyond the scope of this thesis. This section presents however an effective and therefore model independent Lagrangian as an extension to the SM that introduces LFV Higgs boson decays and discusses the indirect limits on LFV Higgs boson decays.

2.2.1 Description of Lepton-Flavor-Violating Higgs Boson Decays

Introducing LFV Higgs boson decays in terms of an effective Lagrangian is done by adding non flavor diagonal couplings between the Higgs field H and leptons. The Lagrangian for these couplings

can be written as

$$\mathcal{L}_{\mathrm{LFV}} = \frac{c_{ij}}{\sqrt{2}} H \bar{\hat{\psi}}_{l_{i_L}} \hat{\psi}_{l_{j_R}}, \qquad i \neq j, \qquad i,j = 1,2,3, \tag{2.30}$$

where c_{ij} is a coupling constant and $\hat{\psi}_{l_i}$ is the spinor of the charged lepton of generation i with flavor f_i in the mass basis, following the notation from section 2.1. Adding $\mathcal{L}_{\mathrm{LFV}}$ to $\mathcal{L}_{\mathrm{SM}}$ does not change the coupling of the Higgs field to same flavor leptons, as it does not add any flavor diagonal couplings ($i = j$), but adds non flavor diagonal couplings that result in LFV decays of the Higgs boson.
In the SM, it was possible to avoid non flavor diagonal couplings between the Higgs field and fermions by changing the fermions base from the flavor into the mass basis. This is possible as the Yukawa couplings introduce the same coupling matrix for the $H\bar{\hat{\psi}}_{l_{i_L}} \hat{\psi}_{l_{j_R}}$ vertex and $v\bar{\hat{\psi}}_{l_{i_L}} \hat{\psi}_{l_{j_R}}$ vertex defining the mass basis, which guarantees that the mass basis diagonalizes the Higgs boson to fermion couplings in flavor space. Please note that the additional terms introduced with $\mathcal{L}_{\mathrm{LFV}}$ modify only the vertex describing Higgs boson lepton coupling and not the leptons' mass terms. This makes it in general impossible to remove the LFV terms introduced by $\mathcal{L}_{\mathrm{LFV}}$ with a change of the fermions' basis, as it is done in the SM, which means that $\mathcal{L}_{\mathrm{LFV}}$ introduces LFV decays of the Higgs boson.

2.2.2 Indirect Limits on Lepton-Flavor-Violating Higgs Boson Decays

Adding a non flavor diagonal coupling between the Higgs boson and leptons has in general more implications than causing LFV Higgs boson decays. The non flavor diagonal couplings will appear for example in higher order corrections in processes containing leptons and change, in absence of cancellations, the rate of the process compared to the SM prediction. Precision measurements of suitable processes allow therefore to infer limits on the non flavor diagonal couplings and thus on the branching ratio of a LFV Higgs boson decay.

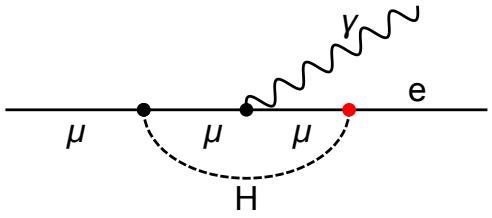


Figure 2.3: Example of the contribution of the LFV Higgs coupling $H\mu e$ (red dot) to the process $\mu \to e\gamma$. Null searches for this process allow to infer limits on $Br(H \to \mu e)$.

Sensitive processes for constraining non flavor diagonal couplings are for example null searches for $\mu \to e\gamma$ [44], $\tau \to e\gamma$ and $\tau \to \mu\gamma$ [45,46]. Figure 2.3 shows a contribution of a LFV Higgs coupling to the process $\mu \to e\gamma$. References [33] and [34] combine the results of those null searches with other sensitive low energy measurements to calculate limits on the couplings c_{ij} of equation 2.30 and therefore on the branching ratio of LFV Higgs boson decays. The null searches for $\mu \to e\gamma$ constrain the branching ratio to $Br(H \to \mu e) < \mathcal{O}(10^{-8})$ leaving, in the absence of cancellations, no hope for observing this decay at the LHC. The constraints on $H \to \tau\mu$ and $H \to \tau e$ are less restrictive. They can have individually a branching ratio of up to $\mathcal{O}(10\%)$, but the strong limit on $\mu \to e\gamma$ does not allow them to be large simultaneously. This means that either $H \to \tau\mu$ or $H \to \tau e$, but not both simultaneously, can have a branching ratio of up to $\mathcal{O}(10\%)$ which is, if realized by nature, well observable at the LHC.

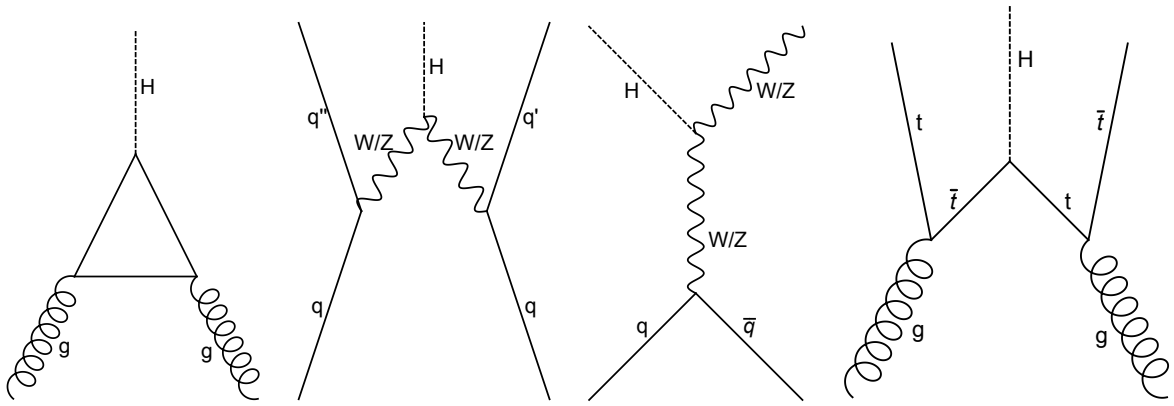
2.3 Phenomenology of the Higgs Boson

There is one unambiguous footprint of the Higgs mechanism in the SM which is the existence of the Higgs boson. Given the mass of the Higgs boson the SM predicts all properties of its Higgs boson. Measuring the properties of the Higgs boson candidate at 125 GeV provides therefore a good possibility to test whether it is in fact the SM Higgs boson, or a Higgs boson of an extended model. The SM does not allow any LFV decays of the Higgs boson, as seen in section 2.1.4. This means observing a LFV Higgs boson decay would be a strong evidence that the Higgs boson candidate is not the SM Higgs boson, but rather the Higgs boson of an extended theory.
This thesis describes the search for LFV decay of the recently discovered Higgs boson candidate. The modeling of LVF Higgs boson decays assumes that the production mechanisms, rates and kinematics are as predicted in the SM, but in addition to the SM decays, also LFV decays are considered. These assumptions are justified, as all measured properties of Higgs boson candidate are in good agreement with the SM prediction. This section gives therefore first a brief overview of the main properties of the Higgs boson followed by a description of LFV Higgs decays and finally a discussion of the experimental discovery of the particle.

2.3.1 Properties of the Higgs Boson

Production of the Higgs Boson

There are four main production processes for the SM Higgs boson at the LHC. Figure 2.4 shows a leading order Feynman graph for each production process ordered from the largest cross section on the left to the smallest cross section on the right.



(a) ggH (b) VBF (c) WH and ZH (d) ttH

Figure 2.4: Examples for leading order Feynman diagrams of the Higgs boson production at the LHC. The diagrams are shown in decreasing order with respect to their cross section (left to right). Please note that the time axis of the shown diagrams runs from bottom to the top and not from the left to the right as in all other shown Feynman diagrams.

The dominant production mode at the LHC is gluon gluon fusion (ggH), as can be seen in table 2.1 that gives the theoretical predicted cross sections for the LHC at a center of mass energy of $\sqrt{s} = 8\,\text{TeV}$. Vector boson fusion (VBF) is, at the LHC, the Higgs boson production mode with the second largest cross section. Although its cross section is more than an order of magnitude

Production mode	ggH	VBF	WH	ZH	ttH
Cross section [pb]	19.27	1.58	0.70	0.42	0.13

Table 2.1: Cross sections of different Higgs boson production processes at the LHC with $\sqrt{s} = 8\,\text{TeV}$ assuming a Higgs boson mass of 125 GeV according to references [47, 48].

smaller than gluon fusion, VBF is still an important production mode at the LHC, as it offers a clear signature in the detector. The two scattered quarks in the final state appear as so called tagging jets in the detector. These tagging jets tend to appear in forward/backward direction with a large separation in η, and the tagging jet system usually has a large invariant mass. Cutting on these properties usually increases the signal to background ratio significantly. Therefore, the search for LFV Higgs boson decays presented in this thesis focuses on the VBF topology.
Higgs strahlung (WH and ZH) has a small cross section compared to VBF with a negligible contribution in the VBF topology, as the only jets present at tree level originate from the decay of the vector boson. This jet system tends to be more central and less separated in η compared to the tagging jet system and therefore rarely passes the VBF selection [4]. The Higgs boson production in association with top quarks (ttH) has a small cross section and is, due to this, negligible in this analysis. This analysis considers therefore only the production modes of ggH and VBF, that have a comparable contribution to the Higgs boson production in the VBF topology, while neglecting all other production modes due to their very small contribution.

Decay of the Standard Models Higgs Boson

The SM Higgs Boson couples to all massive particles and is therefore able to decay at tree level in pairs (particle and anti-particle) of all massive particles. Furthermore, it is able to decay into pairs of massless particles like gluons or photons in loop induced processes. This allows a wide range of

$\boldsymbol{H \to}$	bb	WW	gg	$\tau\tau$	cc	ZZ	$\gamma\gamma$	$Z\gamma$	$\mu\mu$
BR [%]	57.7	21.5	8.6	6.3	2.9	2.6	0.23	0.15	0.02

Table 2.2: Branching ratio (BR) of the SM Higgs boson for a selection of decays assuming a Higgs boson mass of 125 GeV according to references [47, 48].

possible decays of the SM Higgs boson. The branching ratio of each of these decays depends on the Higgs boson mass. Table 2.2 gives the branching ratios for a selection of possible decays assuming a SM Higgs boson with a mass of 125 GeV.
The $H \to \tau\tau$ decay channel of the SM Higgs boson is of special interest for this analysis, as it can produce a very similar final state to the considered LFV Higgs boson decays. Hence, it is an important background in the search for these LFV Higgs boson decays. The branching ratio of the $H \to \tau\tau$ decay is 6.3 %, but only about 12 % of these decays contribute to a fully leptonic final state considered in this analysis due to the hadronic decay modes of the τ lepton. The decays $H \to \mu\mu$ and $H \to ee$ are strongly suppressed compared to the decay $H \to \tau\tau$, since the Higgs boson favors heavy fermions in its couplings (as seen in section 2.1.3) and hence in its decays. Their contribution is therefore negligible within the presented analysis.

[4] The WH and ZH production modes of the $H \to \tau e$ sample combine for example to a total of 0.083 expected events in the signal region of the ee final state, where 3.0 and 4.5 events are expected for the ggH and VBF production mode, respectively.

Lepton-Flavor-Violating Higgs Boson Decays

There is a total of three different LFV Higgs boson decays[5], namely $H \to \tau e$, $H \to \tau\mu$ and $H \to \mu e$. As seen in section 2.2.2, there are very strict indirect limits on $H \to \mu e$, which make a direct observation of this decay at the LHC very unlikely. The analysis presented in this thesis focuses therefore on $H \to \tau e$ and $H \to \tau\mu$. Both decays are assumed to have a branching ratio of 1% in all presented figures and in the calculation of the expected event yields throughout the entire thesis if not stated differently. A branching ratio of 1% is well within the indirect limits presented in section 2.2.2 and of the same order of magnitude as the best fit value for the branching ratio of the decay $H \to \tau\mu$ measured by the CMS collaboration [35].

Figure 2.5: Comparison of schematic graphs of the SM $H \to \tau\tau \to e\mu + 4\nu$ decay (left) and the LFV $H \to \tau e \to e\mu + 2\nu$ decay (right) introducing the labels l_H and l_τ for the charged leptons in the LFV Higgs boson decay. Both decays yield a similar final state with an electron, a muon and missing transverse energy.

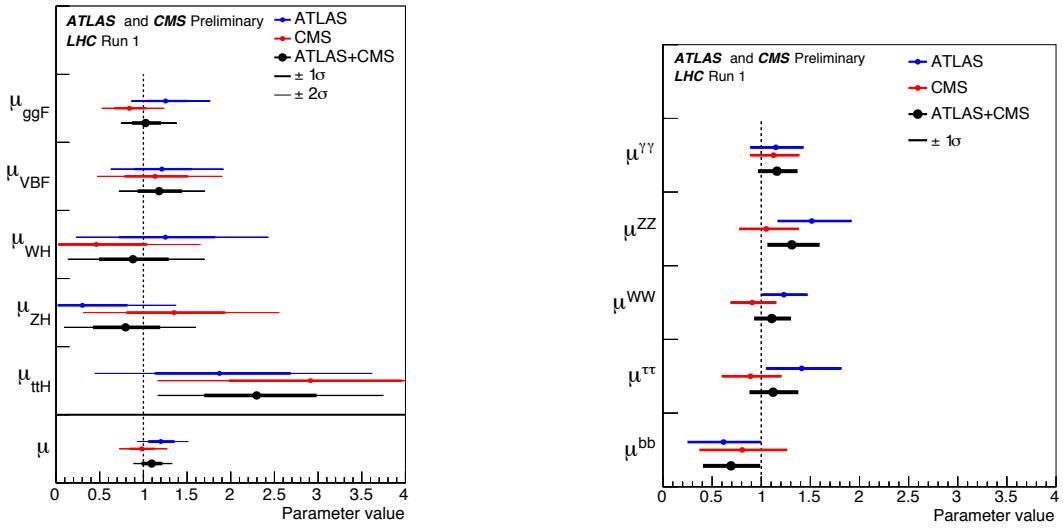
Figure 2.5 shows a comparison of schematic graphs of the SM $H \to \tau\tau$ decay and the LFV $H \to \tau e$ decay in the $e\mu$ final state. From the experimental point of view both decays produce a very similar final state in the ATLAS detector, which is an electron, a neutrino and missing transverse energy (E^T_{miss}) due to the energy carried away by the neutrinos. The τ leptons of both decays are boosted heavily due to the huge mass difference of the Higgs boson and the τ lepton meaning that the momenta of all τ leptons' decay products tend to be parallel to the initial τ leptons' momenta.
Kinematic properties of the charged leptons allow nevertheless to discriminate between both decays. It can be seen easily that the two leptons of the SM $H \to \tau\tau$ decay share the same kinematic properties, while the two leptons of the LFV $H \to \tau e$ are fundamentally different: the lepton l_H (here electron or muon) originating directly from the Higgs boson decay receives on average a total energy corresponding to one half of the Higgs boson mass, while the charged lepton originating from the τ decay receives on average a total energy of approximately one sixth of the Higgs bosons mass. This difference in the leptons' energy yields most likely a measurable difference in the transverse momenta of both leptons, as the Higgs boson tends to receive a negligible boost in its production. There is furthermore an angular correlation between l_τ and the E^T_{miss} in the LFV Higgs boson decays, as only the decay branch of l_τ includes neutrinos. This, and the tendency of all τ lepton decay products to be parallel means that the E^T_{miss} tends to be parallel to l_τ. This does not hold true for the SM $H \to \tau\tau$ decay, as the total E^T_{miss} of this event is given by the vectorial sum of the invisible τ decay products of both decay branches. These fundamentally different properties of the SM $H \to \tau\tau$ and the LFV decay are exploited in the analysis presented in this thesis.

2.3.2 Properties of the Discovered Higgs Boson Candidate

ATLAS and CMS announced in July of 2012 the discovery of a new boson in the mass vicinity of 125 GeV [10, 11]. This boson is a promising candidate for the SM Higgs boson and great effort

[5] Only considering charged leptons.

has been put into the determination of its properties. So far, no significant differences between the Higgs boson candidate's properties and the SM predictions have been observed. The following gives a brief overview on the most important measured properties of the Higgs boson candidate.
The SM predicts all properties of its Higgs boson once the mass of the Higgs boson is known yielding that a precise measurement of the Higgs boson mass is crucial for the investigation of its properties. The mass been measured in a combined analysis of ATLAS and CMS in [49]. The best-fit mass value was found to be $m_H = 125.09 \pm 0.21(\text{stat}) \pm 0.11(\text{syst})\,\text{GeV}$.



(a) Higgs production

(b) Higgs decay

Figure 2.6: Best-fit results for signal strength μ of the different production (a) and decay (b) processes of the Higgs boson candidate normalized to their SM predictions taken from reference [50].

The Higgs boson is the only elementary particle in the SM with spin 0. Hence, the measurement of its spin quantum number provides a good test for the nature of the candidate. Measurements of the Higgs boson candidate's spin have been carried out by ATLAS [51] and CMS [52] favoring the SM hypothesis of a spin 0 with a positive parity.
The measurement of production and decay rates allow additional test of the Higgs boson candidate. These measurements were presented recently in reference [50] combining analyses from ATLAS [53] and CMS [54]. Figure 2.6 shows the best-fit signal strength μ for various production and decay modes of the Higgs boson candidate normalized to the SM prediction. It can be seen that most measured properties are individually in good agreement with the SM predictions. Figure 2.6(a) shows furthermore the best-fit for the global signal strength of the production and decay of the Higgs boson candidate. This is measured to be $\mu = 1.09^{+0.11}_{-0.10}$ deviating by less than 1σ from the SM prediction.
The searches for the decay $H \to \tau\tau$ of the ATLAS [37] and CMS [55] collaboration show furthermore excesses with an observed (expected) significance of 4.4 (3.3) and 3.4 (3.7), respectively. The observed signal strength of this decay is in good agreement with the predictions of the SM. This corresponds to the first evidence for the Yukawa coupling of the Higgs boson to fermions.

Chapter 3

ATLAS Experiment

The ATLAS (**A** **T**oroidal **L**HC **A**pparatu**S**) [56] experiment is one of the two multi-purpose experiments at the LHC (**L**arge **H**adron **C**ollider) [57] at CERN. It is designed to cover a varied

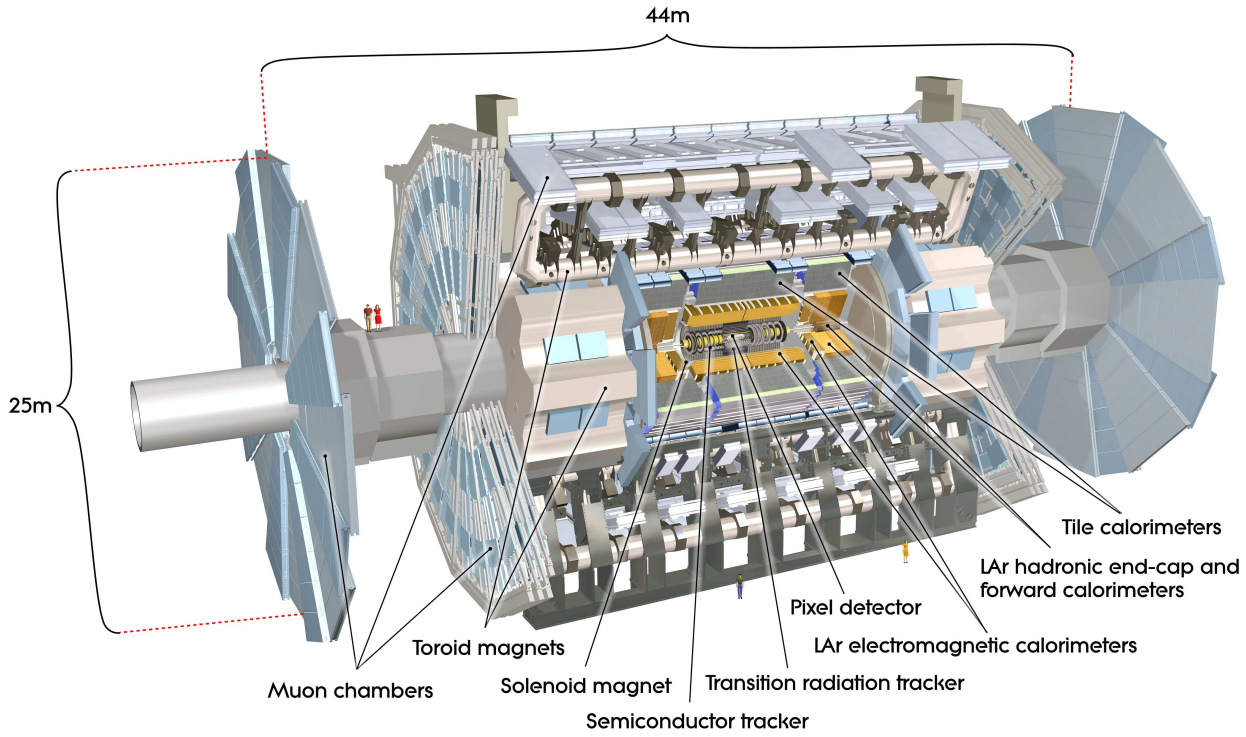


Figure 3.1: Sketch of the ATLAS detector taken from reference [56].

physics program ranging from the precise measurement of the properties of the SM, to searches for physics beyond the SM. Figure 3.1 shows a sketch of the ATLAS detector introducing its three main components that are nested into each other. The innermost component is the inner detector that is dedicated to the measurement of the tracks of charged particles. The calorimeters form the next component providing measurements of the energies of electrons, photons an jets. The muon system is the most exterior component of the ATLAS detector and is dedicated to the measurement of muons.
This chapter defines first the coordinate system used in the ATLAS detector, followed by a brief

description of the three main functional unit of the ATLAS detector. Finally, a short introduction to the trigger system is given. More detailed information on the ATLAS detector can be found in reference [56] on which this chapter is based.

3.1 The ATLAS Coordinate System

ATLAS uses a right-handed coordinate system. The origin of the ATLAS coordinate system is set to the nominal interaction point and the z-axis is defined parallel to the beam direction meaning that the transverse plane coincides with the plane spanned by the x- and y-axis. Furthermore, the x-axis is defined by the straight connection of the nominal interaction point and the center of the LHC ring such that the positive x-axis points towards this center. Finally, the y-axis is selected to point upwards.
The azimuthal angel ϕ and the polar angle θ are introduced in the usual way: measured around and from the z-axis, respectively. The polar angle is used furthermore to calculate the pseudorapidity η defined by

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

The distance ΔR in the plane spanned by ϕ and η is defined as

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

3.2 Inner Detector

The whole inner detector is embedded in the central solenoid creating a magnetic field of 2 T to bend the trajectories of charged particles. The main purpose of the inner detector is to extract information on the tracks of charged particles hence providing on one hand a measurement of their momentum and charge and on the other hand a precise measurement of the primary vertex and potential secondary vertices. It is composed of the pixel and silicon microstrip trackers and the

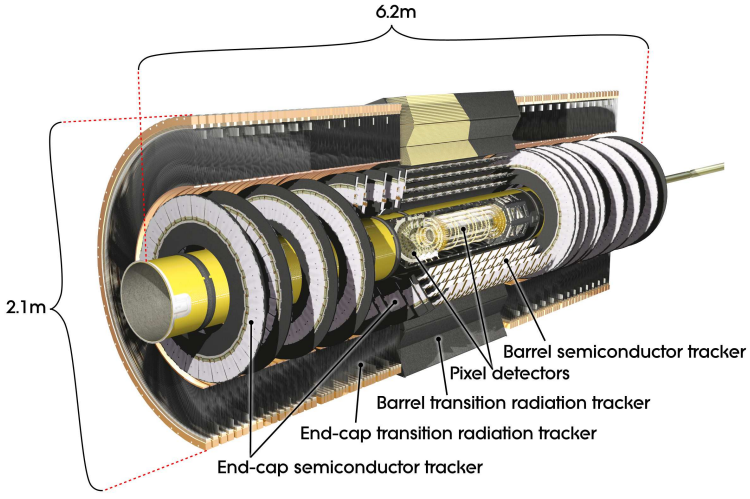


Figure 3.2: Sketch of the ATLAS inner detector taken from reference [56].

straw tubes of the Transition Radiation Trackers. The inner detector is constructed such that it provides the highest granularity in the vicinity of the nominal interaction point. Figure 3.2 shows

a sketch of the inner detector and its main subsystems.
The pixel and silicon microstrip trackers cover the η-region of $|\eta| < 2.5$, while the Transition Radiation Trackers cover only the η-region of $|\eta| < 2$. The required design resolution of the transverse momentum P^T is given by

$$\frac{\sigma_{P^T}}{P^T} = 0.05\,\% \times P^T \oplus 1\,\%, \tag{3.3}$$

using the uncertainty on the P^T measurement denoted σ_{P^T}. Furthermore, the design resolution of the impact parameter $d0$ in the transverse plane is P^T and η dependent. It follows the function

$$\sigma_{d0} = \sigma_{d0}(\infty) \times \left(1 \oplus \frac{P_{d0}}{P^T}\right), \tag{3.4}$$

where $\sigma_{d0}(\infty)$ and P_{d0} are η dependent constants. They take e.g. the values $\sigma_{d0}(\infty) = 10\,\mu$m and $P_{d0} = 14\,$GeV in the η-region of $0.25 < |\eta| < 0.5$.

3.3 Calorimeters

The calorimeters of the ATLAS detector are designed to absorb all detectable particles, except the muons, to measure their energies and extract information on the absorbed particles from the position and topology of the energy depositions. The calorimeters cover a η-range of $|\eta| < 4.9$ and can be divided into two functional units: the electromagnetic and hadronic colorimeters. Figure 3.3 shows a sketch of the ATLAS calorimeter system and its placement around the inner detector.

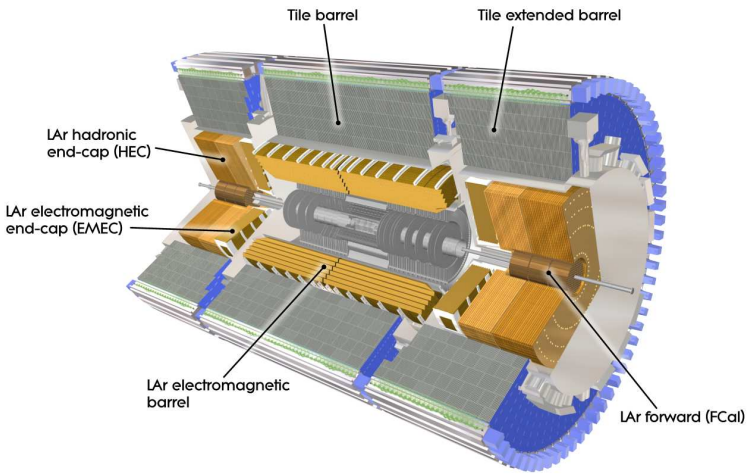


Figure 3.3: Sketch of the ATLAS Calorimeters taken from reference [56].

The electromagnetic calorimeters consist out of lead/liquid-argon (LAr) detectors and are located around the inner detector. Their main purpose is the precise measurement of electrons and photons. The accordion geometry of the electromagnetic calorimeters allows a full coverage of the azimuthal angle ϕ without having any blind spots, while the fine granularity in the central region allows a precise measurement of these objects. The electromagnetic calorimeters are designed to allow a measurement of the energy of the stopped particles in an η-region of $|\eta| < 3.2$ with an energy E resolution of

$$\frac{\sigma_E}{E} = \frac{10\,\%}{\sqrt{E}} \oplus 0.7\,\%. \tag{3.5}$$

The information gathered by the electromagnetic calorimeter in the range of $|\eta| < 2.5$ is used by the trigger system of the ATLAS detector.
The hadronic calorimeters form the second functional layer of the ATLAS calorimeters and are designed to measure the properties of jets. They make use of two types of detectors. The Tile calorimeters are sampling calorimeters using scintillating tiles as the active material and steel as the absorbing material, while the forward and end-cap calorimeters are lead-LAr calorimeters as the electromagnetic calorimeters. The design energy resolution of the hadronic calorimeters in the barrel and end-cup is

$$\frac{\sigma_E}{E} = \frac{50\,\%}{\sqrt{E}} \oplus 3\,\% \tag{3.6}$$

in the η-range of $|\eta| < 3.2$, while the forward calorimeters provide a measurement of the energy in the range of $3.1 < |\eta| < 4.9$ with a resolution of

$$\frac{\sigma_E}{E} = \frac{100\,\%}{\sqrt{E}} \oplus 10\,\%. \tag{3.7}$$

The trigger system makes use of the information of the hadronic calorimeter in the complete measured η-range.
The resolution of the missing transverse energy E^T_{miss} can be parametrized as a function of the total transverse energy measured by the calorimeters, denoted by$\sum E^T$. It is given by

$$\sigma_{E^T_{\text{miss}}} = \alpha \times \sqrt{\sum E^T}. \tag{3.8}$$

The expected values for the constant α range from 0.53 to 0.57.

3.4 Muon System

Muons are the only detectable particles that are capable to pass the calorimeters, hence it is not possible to measure their energy properly with the calorimeter system. They are therefore measured in the muon system forming the exterior component of the ATLAS detector. Figure 3.4 shows a sketch of this muon system.
The muon system extracts the properties of the muons by tracking their trajectories precisely. This is possible due to the deflection of the muon trajectories in a magnetic field. This magnetic field is provided by the large superconducting air-core toroid magnets in the η-region of $|\eta| < 1.4$ and the large barrel toroid in the η-region of $1.4 < \eta < 2.7$.
Monitored Drift Tubes provide a precise measurement of the muon tracks in the η-region of $|\eta| < 2.7$ ($|\eta| < 2.0$ at the innermost layer). Cathode Strip Chambers are additionally used for tracking in the η-region of $2.0 < |\eta| < 2.7$. Resistive Plate Chambers ($|\eta| < 1.05$) and Thin Gap Chambers ($1.05 < |\eta| < 2.7$) are used for triggering and an additional measurement of the muons' tracks. The designed resolution of the transverse momentum P^T by the muon system is given by

$$\frac{\sigma_{P^T}}{p^T} = 10\,\% \tag{3.9}$$

for a muon with a P^T of 1 TeV. The muon system provides a measurement of the muons in the η-range of $|\eta| < 2.7$ and information from the η-region $|\eta| < 2.4$ is used for triggering.

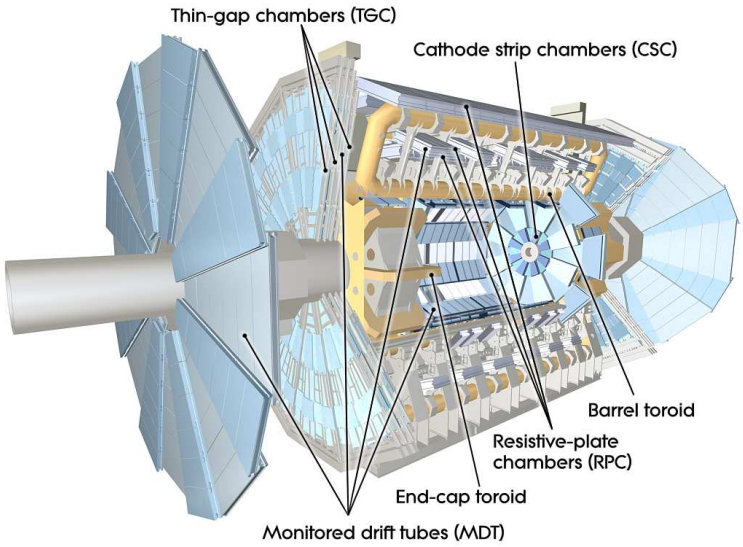


Figure 3.4: Sketch of the ATLAS muon system taken from reference [56].

3.5 Trigger System

The LHC delivers, when running at its design settings, a bunch crossing rate of 40 MHz. In contrast to this high rate, the ATLAS detector is only capable to write events with a rate of $\mathcal{O}(100\,\text{Hz})$ to storage. Therefore, it is important to have an efficient and reliable trigger system that selects events for recording to ensure that the most interesting (in terms of the intended physical analyses) events will be recorded. The trigger system [58] of ATLAS detector decides on three consecutive levels if events are selected for recording. These levels of the trigger system, L1, L2 and event filter, will be described briefly in the following.

The hardware based trigger L1 makes its decision on keeping an event based on the information of a subset of the detector systems. It selects events that contain either high transverse momentum muons, electrons, photons, jets or hadronically decaying τ leptons or events that show a large missing or total transverse momentum. It defines additionally Regions-of-Interest (RoI) given by the regions in the η-ϕ-plane, where its selection process has identified interesting objects and passes this information to the L2 trigger. The L1 trigger makes that decision in less than 2.5 μs and reduces the rate to approximately 75 kHz.

The L2 trigger and the event filter are, in contrast to L1, software based triggers. The selection of L2 is based on the RoI selected in L1. It can make use of the full detector information in these RoIs to decide whether one of the events selected by L1 is kept or discarded by enforcing e.g. tighter P^T-cuts or isolation criteria on the now more precisely measured objects. L2 has an event processing time of roughly 40 ms and reduces the rate to approximately 3.5 kHz. The event rate is after L2 low enough to reconstruct the full event in the event filter. This uses then information of the fully reconstructed event to trim the event rate to a recordable level of 400 Hz, with an average processing time of 4 s.

Chapter 4

Signal and Background Processes

This chapter introduces the signal and background processes considered in this analysis by explaining how the considered background processes contribute to the background of the analysis and by explaining their estimates (via MC simulation or via a data driven approach). This analysis is focused on the VBF topology and therefore accepts only events with at least two jets in the event selection described in section 5.3. This description of the signal and background processes given in this chapter is hence focused on processes that include already two particles in the final state that can be detected as jets. The given Feynman diagrams are examples for possible contributions to a given background illustrating how this background is able to produce a similar final state as the signal and do neither demand to give a complete picture of all possible processes contributing to the background, nor demand to show the diagram with the biggest contribution to the background.

4.1 Signal Processes

The signal samples for the LFV decays of the Higgs boson are produced with NLO accuracy using the POWHEG [59–62] event generator interfaced with PYTHIA8 [63] to provide the parton showering, hadronization and the modeling of the underlying event that is simulated using the AU2 [64] parameter tuning of PYTHIA8. The CT10 [65] parametrization of the parton distribution functions is used. Additionally, the LFV Higgs boson decays are modeled with EvtGen [66].
These signal samples are inclusive samples meaning that they include all possible decays of the τ lepton. Their cross section is set for this analysis to 1 % of the cross section of the SM Higgs boson. This corresponds to assuming a branching ratio of the considered decays of 1 %. Table 4.1 lists all considered signal processes, the used MC event generators and the assumed cross section and the number of simulated MC events.

Process	Generator	Cross section [pb]	# simulated events
ggH $H \to \tau\mu$	POWHEG + PYTHIA8 + EvtGen	0.1927	250000
VBF $H \to \tau\mu$	POWHEG + PYTHIA8 + EvtGen	0.01578	250000
ggH $H \to \tau e$	POWHEG + PYTHIA8 + EvtGen	0.1927	250000
VBF $H \to \tau e$	POWHEG + PYTHIA8 + EvtGen	0.01578	250000

Table 4.1: List of the used signal samples, the used MC generators, the assumed cross section and the number (#) of simulated events. The cross section corresponds to a branching ratio of 1 % for the considered decays of a Higgs boson with a mass of 125 GeV.

This analysis considers only fully leptonic decays of the signal produced in the VBF topology. Figure 4.1 gives an example of a leading order Feynman diagram of a fully leptonic LFV decay of the Higgs boson that was produced in the VBF production mode. It shows that the final state consists of six particles: two opposite charged leptons l_H and l_τ, two neutrinos ν and two quarks. The flavor of l_H is fixed by the considered decay, while l_τ can be either an electron or a muon, meaning that the signal can produce a final state with two leptons of same or of different flavor.
The reconstructed final state is characterized by two charged leptons of opposite sign, two forward jets from the scattered quarks and missing transverse energy E^T_{miss} caused by the neutrinos that cannot be detected directly. The two charged leptons and the E^T_{miss} in the final state have in the signal process distinct properties, e.g. a strong angular correlation between E^T_{miss} and l_τ, that is used in this analysis to discriminate signal and background. A detailed description of these properties is given in section 2.3.1.

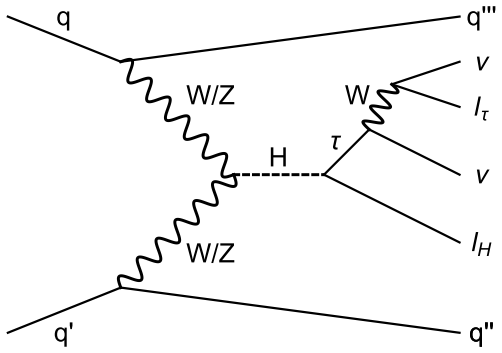


Figure 4.1: Example of a leading order Feynman diagram of the production and fully leptonic LFV decay of the Higgs boson produced in the VBF production mode.

The advantage of the restriction of the analysis to the VBF topology is the presence of the two jets, denoted as the tagging jets, in the final state. The properties of the tagging jets allow to infer information on the production of the Higgs boson. This additional information is independent of the information obtained from the decay products of the Higgs boson, allowing for an additional possibility to discriminate signal and background.
The two tagging jets tend to have a large separation in η and a large invariant mass, as discussed in section 2.3.1.
The production mode of ggH does not yield any jets in the final state at tree level meaning that only higher order processes of the gluon fusion will contain the necessary jets to be selected in the VBF topology. The relative contribution of the ggH production mode is hence suppressed in the VBF topology. Figure 4.2 gives a possible Feynman diagram that yields a final state in the detector that is similar to the VBF topology. The additional two jets that are originating from the two gluons in the given example, will nevertheless differ most likely in their properties from the tagging-jets in the VBF production mode, since they were produced by a different mechanism.

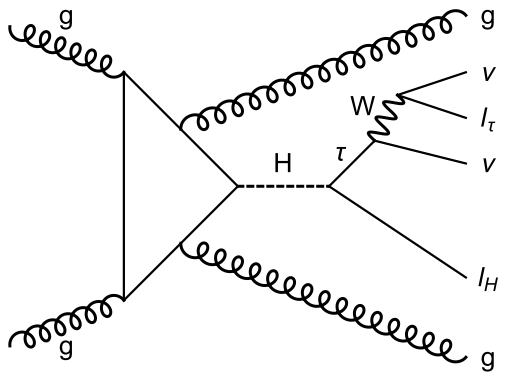


Figure 4.2: Example of a higher order Feynman diagram of the production and fully leptonic LFV decay of the Higgs boson produced in the ggH production mode yielding a similar final state as the same decay of a Higgs boson that was produced in the VBF production mode.

4.2 Background Processes

All background estimates are based on MC simulations, except the $Z \to \tau\tau$ and fake-lepton background. The fake-lepton background describes the contribution of events that contain at least one reconstructed lepton (μ or e) that does not correspond to a physical leptons. This section discusses at first the background taken from MC simulations and gives then an explanation of the partially data driven estimate of the $Z \to \tau\tau$ background and the fully data driven estimate of the fake-lepton background.

4.2.1 MC Based Background Estimates

This section discusses first why each of the backgrounds obtained from MC calculation contributes to the background of this analysis. This is followed by an overview on the MC simulations of these backgrounds.

Top Quark Background

The top quark background[1] considers the single top quark production and the top quark pair production. Both can yield to a final state very similar to the signal. Figure 4.3 shows an example of a leading order Feynman diagram of the top quark pair production and an example of a next to leading order Feynman diagram of the single top production in the Wt-channel with fully leptonic decays of the $W^\pm$ boson. Both yield a final state that is very similar to the signal, as it contains two leptons of opposite charge, two jets originating from the quarks and gluons in the final state and missing transverse energy caused by the presence of neutrinos. Both processes can produce two same flavor of two different flavor leptons, due to the independent decays of the two $W^\pm$ bosons. It is nevertheless possible to discriminate between the signal and the top quark background. This background shows in general not the same correlations between E^T_{miss} and the signal which yields a good separation between signal and background in variables exploiting these correlation, like the reconstructed Higgs boson mass. Furthermore, the present jets are produced by a different mechanism in the VBF signal and the top background allowing to separate both by the properties

[1] Please note: *Top* is used as abbreviation for the top quark background in the legend of some figures.

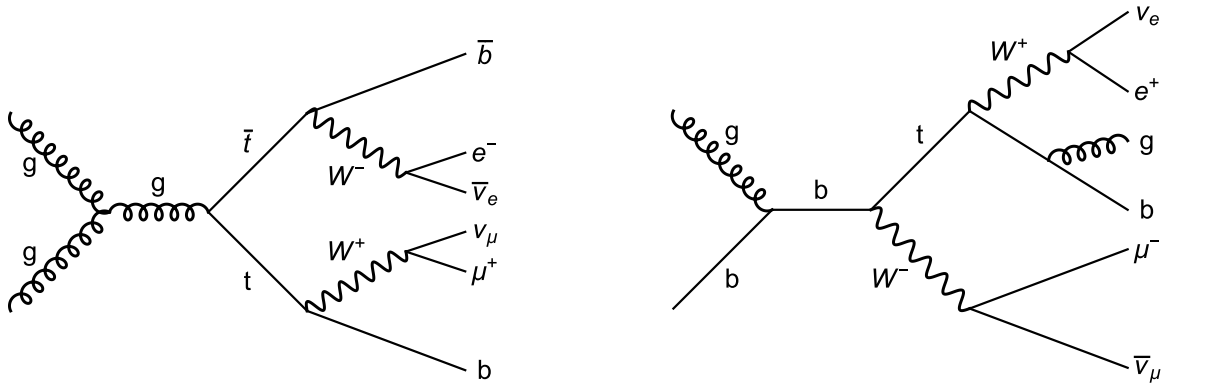


Figure 4.3: Example of Feynman diagrams of the top quark pair (left) and single top quark (right) production yielding a final state similar to the signal.

of their jets.
The top quark decays furthermore almost exclusively into a b quark and a W boson yielding that all processes of the top quark background have most likely a b quark in their final state. The signal does not produce any b quarks at tree level making it very unlikely that a signal event contains a b quark in its final state. This allows to suppress the contribution of the top quark background additionally via a veto against b-tagged jets.
It is nevertheless an important background in this analysis, since the identification of the b-jets has only an efficiency of about 80 % and the background has a large cross section compared to the signal, which can be seen in table 4.2 and 4.1. This background is normalized to data using a dedicated control region, if available. Otherwise it is normalized according to the SM prediction applying systematic uncertainties on the cross section. The data driven normalization procedure is described in section 7.4.1.

Di-Boson Background $WW/WZ/ZZ$

This background includes the production of all combinations of two W and Z bosons. Figure 4.4 shows three possible Feynman diagrams that produce a final state that is similar to the final state of the signal in the VBF production mode. The processes described by the Feynman diagrams given in figure 4.4 (a) and (c) yield at tree level a final state with two same flavor leptons of opposite charge and two jets originating from the hadronic decay of the W or Z boson, but no objects that cause E^T_{miss}. It is nevertheless possible that some E^T_{miss} is associated with them due to a mismeasurement of E^T_{miss}. The higher order diagram given in 4.4 (b) yields a final state with two jets, two same or different flavor leptons of opposite charge and missing transverse energy due to the neutrinos.
This discussed diagrams are just a subset of possible diagrams leading to a final state similar to the one of the VBF signal, but demonstrate nevertheless that the $WW/WZ/ZZ$ background can produce such a final state with same and different flavor leptons of opposite charge in many different ways, some with two jets at tree level and some with a physical source for E^T_{miss}. This and its relatively large cross section makes this background important for the presented analysis.
All processes that contribute to this background have in common that the leptons and E^T_{miss} and the jets have in general not the same properties as the signal, since they are produced by different processes. The invariant mass of the jets originating from the decay of a Z or W boson corresponds to the reconstructed mass of the boson and is therefore relatively small compared to the invariant

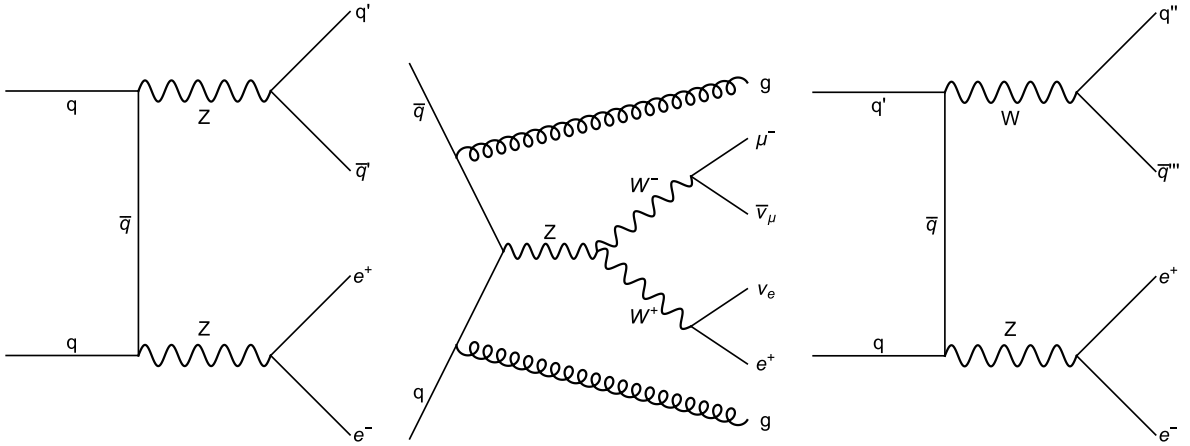


Figure 4.4: Example of Feynman diagrams of the ZZ (left), WW (middle) and WZ (right) production yielding a final state similar to the signal.

mass of the two tagging-jets in the VBF signal events. The leptons produced by this process differ for example in their correlation with E^T_{miss} from the leptons of signal events, since the E^T_{miss} is in general for the $WW/WZ/ZZ$ processes not strictly linked to one of the leptons, as it is in signal events. This allows nevertheless to separate this background from the signal.

$Z \to \mu\mu$ and $Z \to ee$ Background

The $Z/\gamma^* \to ee$ and $Z/\gamma^* \to \mu\mu$ backgrounds, that will be denoted as $Z \to ee$ and $Z \to \mu\mu$ from here on, are very similar and will be discussed simultaneously. They have by far the largest cross section from all processes considered in this analysis, as can be seen in table 4.2. They do not

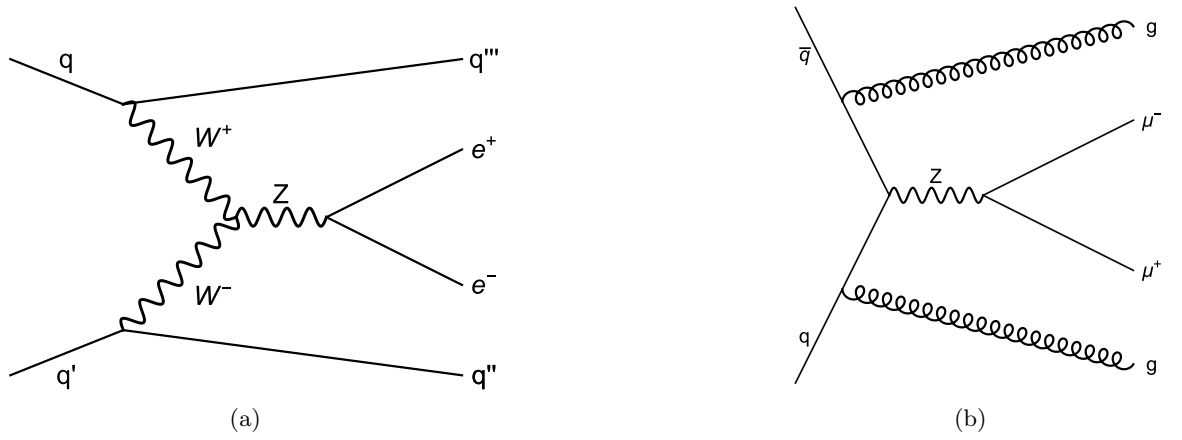


Figure 4.5: Example of Feynman diagrams of the $Z \to ee$ and $Z \to \mu\mu$ production associated with two jets yielding a final state similar to the signal.

contain any jets at tree level, but they have such a large cross section that their production with two associated partons, leading most likely to two jets in the final state, is $\mathcal{O}(10^5)$ more likely than

the production of the VBF signal. Figure 4.5 gives two higher order processes yielding a final state with two same flavor leptons of opposite charge and two jets. The diagram on the left is produced in a VBF structure analogously to the VBF signal yielding the same kinematics of the jets for both processes. A process like the one given in the right diagram is however much more likely since it contains less vertices of the weak interaction, but yields most likely other jet kinematics than the VBF signal. This background contains up to higher order corrections no physical source of E^T_{miss}, but a mismeasurement of E^T_{miss} can nevertheless associate a E^T_{miss} for this background.
The $Z \to ee$ and $Z \to \mu\mu$ background can, in the absence of detector effects, not produce a final state with exactly two different flavor leptons ($= e, \mu$), while the signal can produce such a final state. This yields that final states including exactly two different flavor leptons are barely affected by this large background making them the most promising ones for the search for the signal.
The properties of the leptons, E^T_{miss} and the jets of this background differ nevertheless significantly from the signal, motivating a search for the signal in the final states with two same flavor leptons despite the overwhelming size of the $Z \to ee$ and $Z \to \mu\mu$ background.
The analysis uses additional *VBF-filtered* samples for these backgrounds as can be seen in table 4.2, giving the used background MC samples. This is necessary as the acceptance probability for a $Z \to ee$ or $Z \to \mu\mu$ event in the VBF topology is relatively small, due to the lack of jets at tree level, meaning that huge inclusive MC samples would be necessary to guarantee a sufficient modeling of these backgrounds in the VBF topology. This is avoided by using samples that are filtered at generator level such that most of the simulated events will most likely pass the VBF selection. An overlap removal is performed between the regular and the *VBF-filtered* samples ensuring that no part of the phase space is considered twice. It was furthermore necessary to apply scale factors for both backgrounds to absorb a tension between data and MC prediction in the modeling of the difference in η of the two leading jets. A systematic uncertainty for those scale factors is applied.
The normalization of the $Z \to ee$ and $Z \to \mu\mu$ background is inferred from dedicated control regions, if available, and otherwise from the MC prediction. The description of the data driven normalization is given in section 7.5.1 and 7.5.2.

SM Higgs Background $H \to \tau\tau$

The SM Higgs boson decay in a pair of τ leptons is the smallest considered background with a cross section of the same order of magnitude as the considered signal. It is nevertheless an important background, as this background is produced by the exact same production mode as the signal and decays into a very similar final state. This can be seen in figure 4.6 giving two examples of Feynman diagrams of the SM Higgs boson production and decay that yield a final state similar to the VBF signal. Since the production mode of this background and the signal is identical, it is only possible to discriminate between the two by the properties of their decay. This is possible due to the different kinematics of both decays that were discussed in detail in section 2.3.1. The different kinematics can be exploited in the reconstruction of the Higgs boson mass tailored for the LFV decays of the Higgs boson described in section 7.2.4 allowing the analysis to discriminate between both decays. The $H \to \tau\tau$ background is normalized to the SM prediction.

Overview on MC Simulation Based Backgrounds

Table 4.2 gives an overview on the MC samples that were used for the estimate of the MC based background estimates. It lists the MC generators used for the simulation of the hard scattering process and the model used for the simulation of the parton showers. The table lists furthermore the value for the cross section times the branching ratio and the order of the corresponding perturbative

Background	Sample	Generator	CS × BR [pb]	
$Z/\gamma^* \to ee$ ($Z/\gamma^* \to \mu\mu$)	$60\,\text{GeV} \leq m_{ll} \leq 2\,\text{TeV}$; $NP = 0$	ALPGEN [67] + PYTHIA8	848.4 (848.6)	NNLO [68, 69]
	$60\,\text{GeV} \leq m_{ll} \leq 2\,\text{TeV}$; $NP = 1$		207.3 (207.4)	
	$60\,\text{GeV} \leq m_{ll} \leq 2\,\text{TeV}$; $NP = 2$		69.5 (69.5)	
	$60\,\text{GeV} \leq m_{ll} \leq 2\,\text{TeV}$; $NP = 3$		18.5 (18.5)	
	$60\,\text{GeV} \leq m_{ll} \leq 2\,\text{TeV}$; $NP = 4$		4.7 (4.7)	
	$60\,\text{GeV} \leq m_{ll} \leq 2\,\text{TeV}$; $NP = 5$		1.5 (1.5)	
	$10\,\text{GeV} \leq m_{ll} \leq 40\,\text{GeV}$; $NP = 0$	ALPGEN + PYTHIA8	699.7 (726.4)	NNLO [68, 69]
	$10\,\text{GeV} \leq m_{ll} \leq 40\,\text{GeV}$; $NP = 1$		51.1 (52.0)	
	$10\,\text{GeV} \leq m_{ll} \leq 40\,\text{GeV}$; $NP = 2$		24.9 (25.4)	
	$10\,\text{GeV} \leq m_{ll} \leq 40\,\text{GeV}$; $NP = 3$		5.8 (5.8)	
	$10\,\text{GeV} \leq m_{ll} \leq 40\,\text{GeV}$; $NP = 4$		1.8 (1.9)	
	$40\,\text{GeV} \leq m_{ll} \leq 60\,\text{GeV}$; $NP = 0$	ALPGEN + PYTHIA8	36.4 (36.7)	NNLO [68, 69]
	$40\,\text{GeV} \leq m_{ll} \leq 60\,\text{GeV}$; $NP = 1$		6.4 (6.4)	
	$40\,\text{GeV} \leq m_{ll} \leq 60\,\text{GeV}$; $NP = 2$		2.2 (2.3)	
	$40\,\text{GeV} \leq m_{ll} \leq 60\,\text{GeV}$; $NP = 3$		0.6 (0.6)	
	$40\,\text{GeV} \leq m_{ll} \leq 60\,\text{GeV}$; $NP = 4$		0.2 (0.2)	
	VBF-filtered; $NP = 2$	SHERPA [70]	3.8 (1.4)	LO [70]
	VBF-filtered; $NP = 3$		2.2 (1.2)	
	VBF-filtered; $NP = 4$		0.9 (0.6)	
	VBF-filtered; $NP = 5$		0.4 (0.4)	
	EW	SHERPA	0.4 (0.4)	
Top-Quark	single Top; s-channel; leptonic decay	POWHEG [59–62] + PYTHIA8	1.8	NNLO [71]
	single Top; t-channel; leptonic decay	AcerMC [72] + PYTHIA6 [73]	28.4	NNLO [74]
	single Top; Wt-channel	POWHEG + PYTHIA8	22.4	NNLO [75]
	$t\bar{t}$	POWHEG + PYTHIA8	129.3	NNLO+NNLL [76–81]
$WW/WZ/ZZ$	ZZ	HERWIG [82]	1.6	NLO [83]
	WZ	HERWIG	6.8	NLO [83]
	$qq \to WW \to ll\nu\nu$; $NP = 0$	ALPGEN + HERWIG	3.0	NLO [83]
	$qq \to WW \to ll\nu\nu$; $NP = 1$		1.5	
	$qq \to WW \to ll\nu\nu$; $NP = 2$		0.7	
	$qq \to WW \to ll\nu\nu$; $NP = 3$		0.4	
	$qq \to WW \to qql\nu$; $NP = 0$	ALPGEN + HERWIG	12.6	NLO [83]
	$qq \to WW \to qql\nu$; $NP = 1$		6.3	
	$qq \to WW \to qql\nu$; $NP = 2$		3.0	
	$qq \to WW \to qql\nu$; $NP = 3$		1.7	
	$gg \to WW \to ll\nu\nu$; $NP = 3$	GG2WW [84] + HERWIG	0.2	NLO [84]
SM $H \to \tau\tau$	ggH (125 GeV): $H \to ll + 4\nu$	POWHEG +PYTHIA8	0.15	NNLO+NNLL [47, 85–93]
	VBF (125 GeV): $H \to ll + 4\nu$	POWHEG +PYTHIA8	0.01	(N)NLO [47, 94–96]

Table 4.2: List of the used background samples, the used MC generators, the cross section (CS) that is multiplied with the corresponding branching ratio (BR) of all non-inclusive samples for all backgrounds modeled by MC simulations. The variable NP denotes the number of additional partons that are included in the matrix element used in the MC simulation. Please note that the $Z/\gamma^* \to ee$ and $Z/\gamma^* \to \mu\mu$ backgrounds are listed together. The CR×BR values given in the brackets correspond to the ones for the $Z/\gamma^* \to \mu\mu$ background.

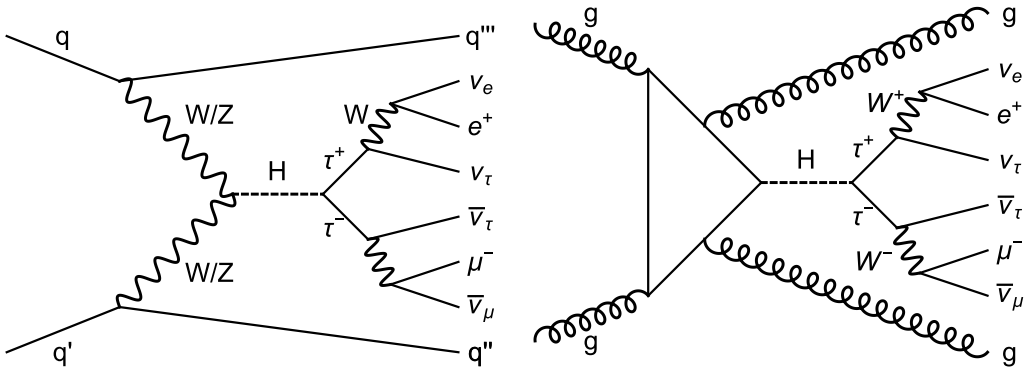


Figure 4.6: Example of a leading order Feynman diagram of the production and fully leptonic decay of the SM Higgs boson via two τ leptons produced in the VBF production mode (left). Higher order Feynman graph of the production and fully leptonic decay via two τ leptons of a SM Higgs boson in association with two gluons (right) yielding a similar final state as the VBF signal process.

QCD calculation.
The ATLAS detector response [97] is simulated with the GENAT4 program [98]. The underlying events are simulated using the AU2 [64] parameter tuning of PYTHIA8. Additionally, the CT10 [65] parametrization of the parton distribution functions is used. It can be seen that the cross section of most backgrounds is by multiple orders of magnitude larger than the assumed cross section of the signal processes that were given in table 4.1, foreshadowing the main challenge of this analysis: the poor signal to background ratio.

4.2.2 $Z \to \tau\tau$ Background Estimate

The $Z \to \tau\tau$ background is an irreducible background in this search for the LFV decays $H \to \tau\mu$ and $H \to \tau e$, since a fully leptonic decay of $Z \to \tau\tau \to ll + 4\nu$ ($l = \mu, e$) produces a similar final state of two leptons ($= \mu, e$) and missing transverse energy in the detector. This can be seen in figure 4.7 showing two possible Feynman diagrams for this background yielding a final state similar to the VBF signal. The differences in the kinematic of the leptons and their correlations to E^T_{miss} exploited e.g. by the mass reconstruction of the Higgs boson allow nevertheless to discriminate well between signal and this background.
The $Z \to \tau\tau$ background is estimated using the mostly data driven embedding procedure. It is based on $Z \to \mu\mu$ data events that are very similar to the $Z \to \tau\tau$ events up to small deviations due to the mass difference of the muon and τ lepton. They can be used as an input for this background estimate, as they can be selected with a high purity and the selected $Z \to \mu\mu$ sample contains almost no Higgs signal due to the small branching ratio of $H \to \mu\mu$.
This background estimate uses the selected $Z \to \mu\mu$ data events, removes the muons from the event and replaces them with simulated $Z \to \tau\tau$ decays creating hence a hybrid event that is partially obtained from simulation and data. A detailed description of this procedure can be found in reference [99]. This procedure has two key advantages compared to a classical fully MC simulation based estimate. The first advantage is that it is less prone to mismodeling, since all parts of the event not describing the $Z \to \tau\tau$ decay are obtained from data. This yields directly the second advantage given by the fact that this background estimate is affected by less systematic

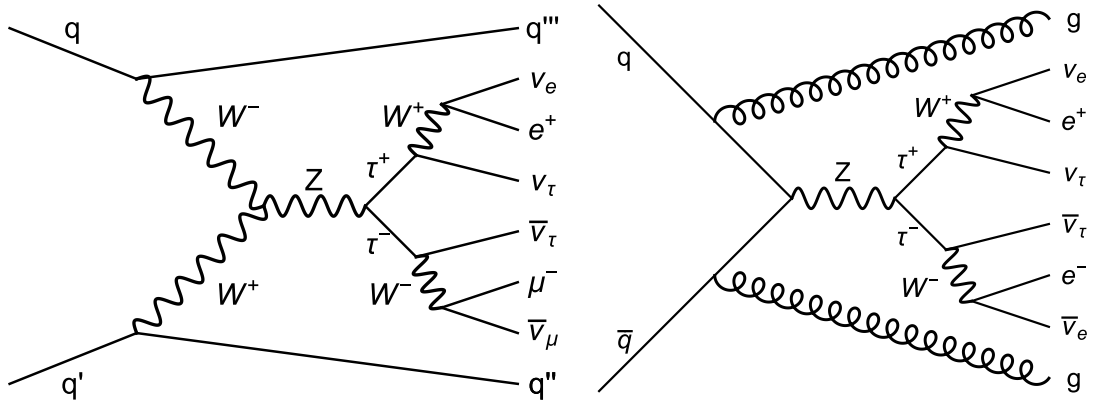


Figure 4.7: Two examples of Feynman diagrams of the production and fully leptonic decay of the Z boson produced in association with jets via two τ leptons. The VBF production process shown in the left diagram yields a final state very similar to the VBF signal.

uncertainties as most components of the event e.g. jets or pile-up are taken from data and not from simulation.
This background estimate does not predict the expected event yield and is therefore normalized to the event yield predicted by the $Z \to \tau\tau$ MC samples. The normalization factor N is given by

$$N = \frac{\text{Event yield embedded sample}}{\text{Event yield } Z \to \tau\tau \text{ MC sample}}. \tag{4.1}$$

This is done for each considered final state introduced in section 7.3.4 separately after Cut9 of the basic VBF selection described in section 5.3. The obtained normalization factor for the embedded sample are listed in table 4.3 for each final state.

Final state	Normalization factor
ee	0.192
$\mu\mu$	0.136
$e\mu$	0.159
μe	0.157

Table 4.3: Normalization factors for the embedded $Z \to \tau\tau$ background sample.

4.2.3 Fake-Lepton Estimate

Fake-leptons[2] (lepton means here e or μ) are reconstructed leptons that do not correspond to physical leptons. They arise e.g. from jets that are misidentified as leptons, which happens rarely, but can nevertheless yield to a significant contribution of the total background given the huge production rate of jets in the ATLAS experiment. Fake-leptons are a background for this analysis as they basically add a lepton to an event, making events that have only one physical lepton pass the event selection. The events with a the fake-lepton can be by chance very signal-like, but this is not very likely resulting mostly in a small contribution of the fake-lepton background to the total

[2]Please note: *Fake* is used as abbreviation for the fake-lepton background in the legend of most figures.

background. It is nevertheless an important background since it is difficult to model, resulting in large uncertainties on the modeling that can decrease the sensitivity of the analysis. The modeling of this backgrounds is very challenging as it is not possible to simulate it at a reasonable precision with MC simulations, as the low misidentification rate would call for gigantic MC samples. Therefore, the fake-lepton background is estimated in a fully data driven way.

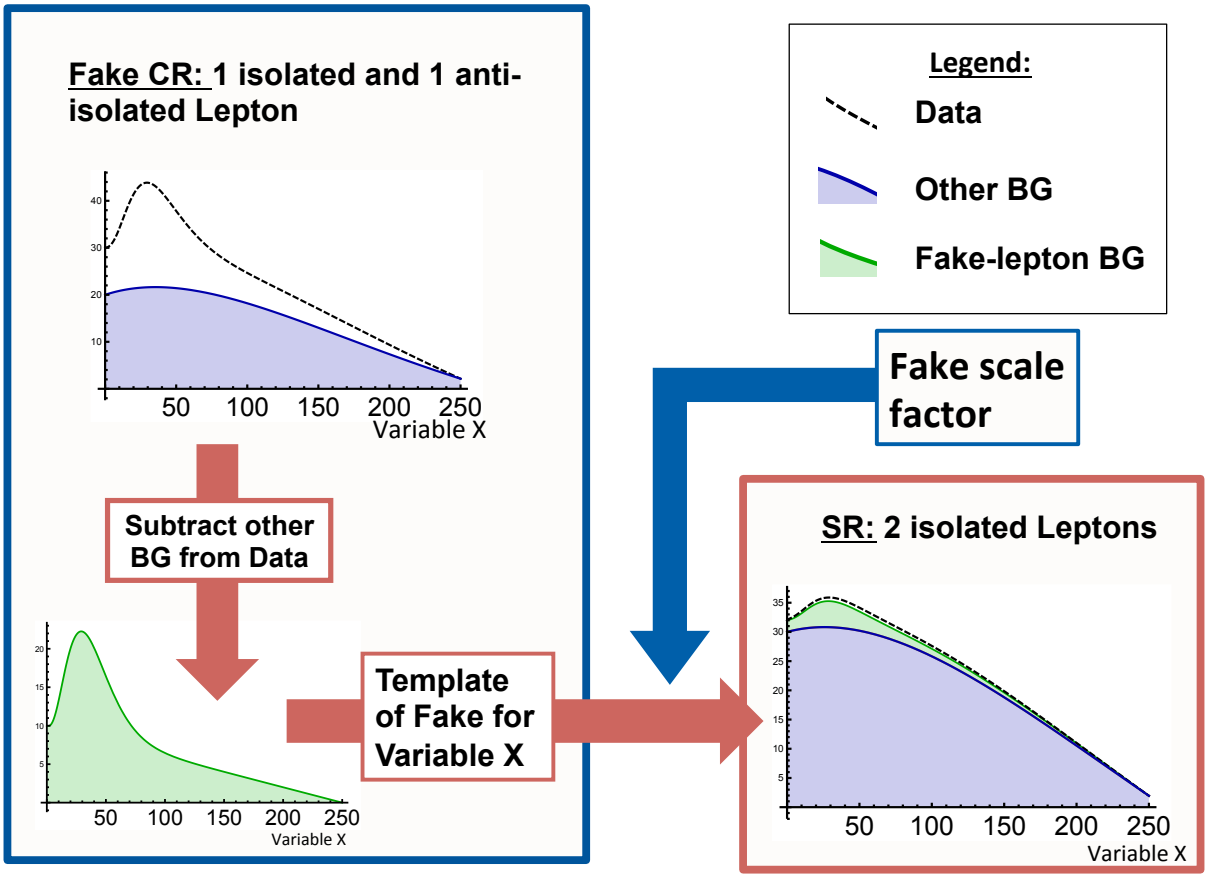


Figure 4.8: Illustration of the estimate of the fake-lepton background. BG is used as an abbreviation for background.

The estimate of the fake-lepton background is a template-based method illustrated in figure 4.8. It uses two different regions: the signal region and the fake control region. The fake control region is defined exactly as the signal region, but the isolation requirement for one of the leptons is reversed yielding that this region is enriched in the fake-leptons that tend to be anti-isolated. The contribution of the fake background in a given variable X is calculated in to steps.
First, the template for the fake-lepton background is calculated. It is derived in the fake control region by subtracting all other background estimates from the observed data in the distribution of the variable X. This template corresponds to the fake-lepton background contribution in the fake control region in the distribution for the variable X. Now it is assumed that the shape of the distribution of the fake-lepton background in the variable X is identical in both region, as they are identical with respect to their kinematics due to the application of the same cuts. Within this

assumption, it is possible to use the template of the fake-lepton background from the fake control region to model the shape of the fake-lepton background in the signal region. This means that the fake-lepton background estimate is fixed up to its normalization.
For the calculation of the normalization it is assumed that the ratio of the event yields of the fake-lepton background between the signal and fake control region is independent of the applied event selection and therefore constant in the cut-flow. This means that the normalization of the fake-lepton background can be calculated by scaling the template with a constant normalization factor, the so called fake scale factor. This fake scale factor is inferred at an early stage of the cut-flow utilizing the smaller statistic uncertainties due to the larger event count. At this early stage the scale factor is variated by a fit in the distribution of a selected variable in the signal region. Eventually, the fake scale factor is selected that yields the smallest χ^2 between the observed date and the sum of all background estimate in the distribution of a selected variable.
This analysis uses the distribution of variable $P_{l\tau}^T$, described in section 7.2. This variable was selected as the distribution of the fake-lepton background peaks at small values of this variable and its contribution decreases rapidly with higher values of this variable yielding stable fit results. A fake scale factor is derived for each of the final states introduced in section 7.3.4. In the different flavor final states the scale factor was determined after Cut9 of the basic VBF selection presented in section 5.3. For the same flavor final states additional requirements had to be introduced in order to avoid potential effects of a mismodeling of the $Z \to \mu\mu$ and $Z \to ee$ background on the inferred fake scale factors. This mismodeling is discussed in more detail in section 7.5.1. The additional requirements and the derived scale factor are listed in table 4.4 for all considered final states.

Final state	Additional cuts	Fake scale factor
ee	$m_{ll} > 30\,\text{GeV}$ and $bTag = 0$	0.22 ± 0.03
$\mu\mu$	$m_{ll} > 30\,\text{GeV}$ and $bTag = 0$	0.337 ± 0.017
$e\mu$	-	0.208 ± 0.002
μe	-	0.212 ± 0.002

Table 4.4: Fake scale factors and the additional cuts that were used on top of Cut1 to Cut9 of section 5.3 for their derivation. The variables used in the additional cuts are given in section 7.2.

Additionally, it was necessary to introduce scale factors for the $Z \to ee$ sample in the ee final state to correct small discrepancies between the background estimates and the the observed data in the $P_{l_\tau}^T$ distributions to receive reasonable results for the fake scale factor in this final state. These scale factors are applied to the $Z \to ee$ sample in the ee final state in the entire analysis and a systematic uncertainty was calculated and considered in the statistical evaluation (see section 8.1.2).

Chapter 5

Physics Object Reconstruction, Used Triggers and Preselection

This chapter discusses the reconstruction of the physics objects, the used triggers and the preselection of events. The first section gives a brief overview on the reconstruction of the physics objects from the recorded response of the ATLAS detector. The next section presents the used triggers followed by the discussion of the event preselection in the last section.

5.1 Physics Object Reconstruction

This section describes the reconstruction of muons, electrons, jets and missing transverse energy from the recorded response of the ATLAS detector. The presented analysis uses the same object reconstructions as in the fully leptonic channels of the SM $H \to \tau\tau$ analysis of the ATLAS collaboration [37].

5.1.1 Muons

The reconstruction algorithm [100] for the muons uses combined information of the inner detector and the muon system. This allows to reconstruct muons in the η-range of $|\eta| < 2.5$, as the inner detector covers only this η-range. Muons are furthermore required to have a transverse momentum of at least 10 GeV[1].
Muons are only accepted if they meet requirements on hits in the inner detector to ensure a precise measurement of the transverse momentum and to reduce the misidentification rate. Additionally, it is required that the distance in the z-position of the reconstructed vertex and the point of closest approach of the muons to the beam-line is less than 1 cm to minimize the contribution of cosmic muons. The reconstruction efficiency of muons fulfilling these criteria is mostly above 95 %. This can be seen in figure 5.1 showing the muon reconstruction efficiency depending on η. The muon reconstruction efficiency drops however in the region of $\eta = 0$ for data due to limitations of the detector. This figure shows additionally small discrepancies between observed data and the MC simulation. Scale factors are applied to correct for them. For more details on the muon reconstruction see reference [100].

[1] This P^T threshold can be raised depending on the used trigger as described in section 5.2.

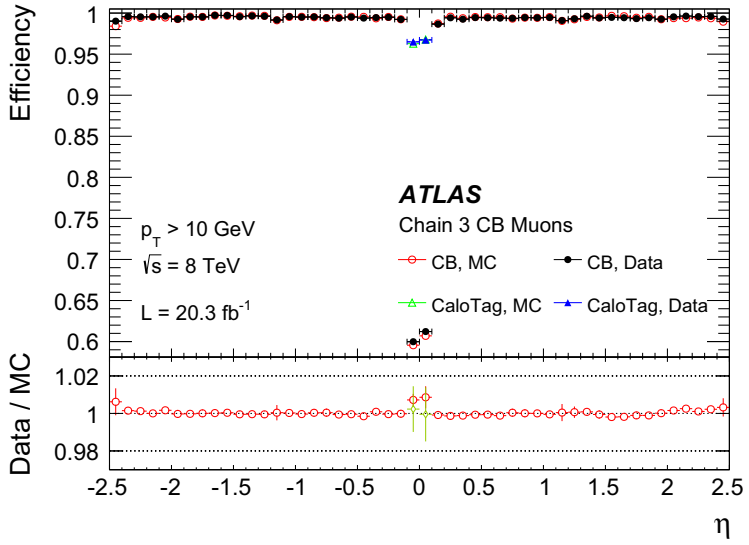


Figure 5.1: Muon efficiency depending on η taken from reference [100]. The muons of this analysis are reconstructed using the combined algorithm denoted as CB in the legend of the figure.

5.1.2 Electrons

This analysis considers only electrons within the pseudorapidity range of $|\eta| < 2.47$. Additionally, electrons are excluded if they were found in the transition region of the end-cap and barrel calorimeters at $1.37 < |\eta| < 1.52$. The reconstruction of electrons [101] is done in two steps, the actual electron reconstruction and the electron identification. A detailed description can be found in reference [101].
The electron reconstruction makes use of information from the inner detector and the electromagnetic calorimeters. First, clusters in the electromagnetic calorimeters are identified. Then tracks from the inner detector are associated to these clusters. If multiple tracks are matched to a cluster the one closest in ΔR is selected, while tracks with hits pixel or silicon microstrip tracker are prioritized. After a successful matching the electron candidate's energy is calculated from the energy deposition in the calorimeter and its estimated energy loss.
The electron identification is realized in a cut based selection defining three types of electrons, *loose*, *medium* and *tight* listed in increasing (decreasing) order of background suppression (identification efficiency). Background refers here to hadrons misidentified as electrons and non isolated electrons originating from semi-leptonic decays of heavy-flavor particles. The definitions of these three electron types can be found in reference [101].
This analyses uses electrons of the type *medium* with a transverse momentum of at least[2] 15 GeV. Combined efficiencies for the reconstruction and identification of electrons depending on η and the transverse energy E^T are given in figure 5.2.

5.1.3 Jets

Jets of this analysis are reconstructed using the anti-k_t algorithm [103,104] in the η-range of $|\eta| < 4.5$ and a minimal transverse momentum of 20 GeV. The anti-k_t clustering algorithm is run with a radius parameter of $R = 0.4$ taking topological energy clusters [105] in the calorimeters as an input.

[2] This P^T threshold can be raised depending on the used trigger as described in section 5.2.

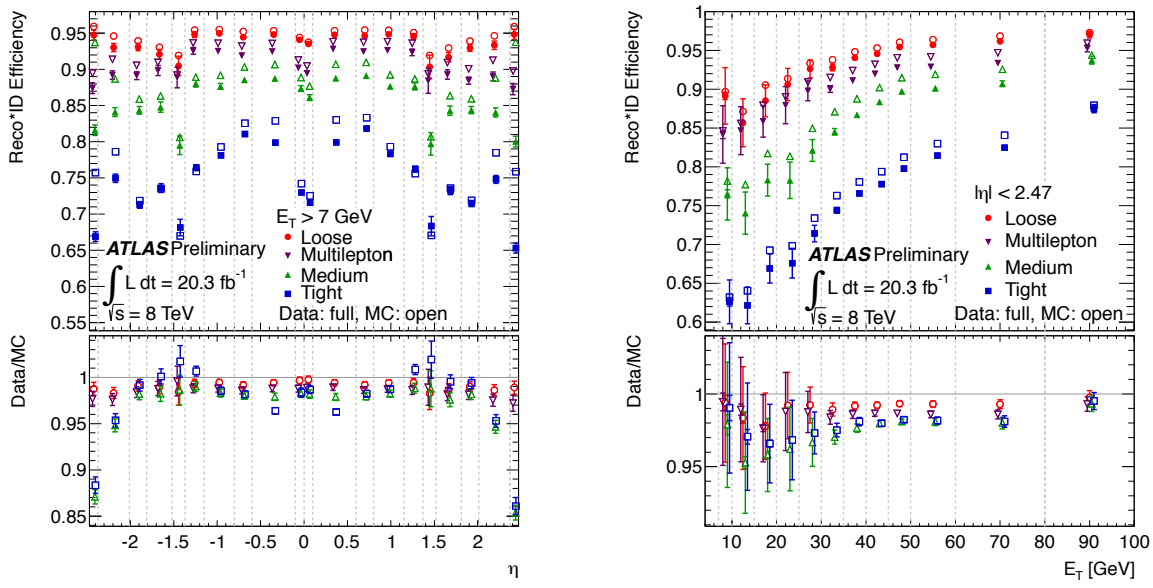


Figure 5.2: Combined reconstruction and identification efficiency for electrons depending on η (left) and E^T (right) taken from reference [102].

A requirement on the jet-vertex-fraction (JVF) of central jets ($|\eta| < 2.4$) is introduced to reduce the contribution of jets originating from additional interactions. The JVF quantifies the likelihood of a jet originating from the primary vertex. It is calculated using the the transverse momenta of the tracks associated to the jet and is defined as

$$\text{JVF} = \frac{\sum P^T \,(\text{tracks pointing towards primary vertex})}{\sum P^T \,(\text{all tracks})}. \tag{5.1}$$

Central jets with $P^T < 50\,\text{GeV}$ are only considered if their JVF is larger than 50 %. This additional requirement is introduced to suppress the contributions of jets that originate from additional interactions happening simultaneously to the considered event.

5.1.4 Missing Transverse Energy

The ATLAS detector is not able to detect neutrinos directly, as they escape the detector without interacting with it. It is nevertheless possible to infer some information on the neutrinos from the properties of all other measured particles, as the protons in the initial state do not have any momentum perpendicular to the beam axis meaning that the initial transverse momentum is zero. This allows to infer information on the neutrinos from the vectorial sum of the transverse momenta of all particles in the final state. The negative of this vectorial sum, denoted as the missing transverse energy $\vec{E}^T_{\text{miss}}$ [106], corresponds in the absence of mismeasurements to the vectorial sum of the transverse momenta of the neutrinos.

The absolute value of $\vec{E}^T_{\text{miss}}$ is defined as

$$E^T_{\text{miss}} = \sqrt{\left(E^T_{\text{miss, x}}\right)^2 + \left(E^T_{\text{miss, y}}\right)^2}, \tag{5.2}$$

using the components of the missing transverse energy in in x and y direction denoted by $E^T_{\text{miss, x}}$ and $E^T_{\text{miss, y}}$. These x- and y-components are calculated using the contributions to E^T_{miss} of all reconstructed physics objects (jets, e, μ, τ_{had} and γ) and the soft energy depositions in the calorimeters, denoted by soft terms, that are not associated to an identified physics object by

$$E^T_{\text{miss, x(y)}} = E^{T,\text{ jets}}_{\text{miss, x(y)}} + E^{T,\,e}_{\text{miss, x(y)}} + E^{T,\,\mu}_{\text{miss, x(y)}} + E^{T,\,\tau_{\text{had}}}_{\text{miss, x(y)}} + E^{T,\,\gamma}_{\text{miss, x(y)}} + E^{T,\text{ soft terms}}_{\text{miss, x(y)}}. \tag{5.3}$$

Furthermore, the individual contributions (cont = jets, e, μ, τ_{had}, γ and soft terms) are calculated by

$$E^{T,\text{ cont}}_{\text{miss, x(y)}} = -\sum E^{T,\text{ cont}}_{\text{x(y)}}. \tag{5.4}$$

Pile-up has a large influence on the measurement of E^T_{miss}. This influence is caused mainly by the soft terms, hence it has to be corrected to limit the effects of pile-up on the E^T_{miss} measurement. A soft term vertex fraction (STVF) is defined in analogy to the JFV given in equation 5.1 and the contributions to E^T_{miss} is scaled by the STVF to correct for the influence of pileup to this term. Figure 5.3 shows the resolution of E^T_{miss} depending on the total transverse momentum without a pile-up suppression and for multiple methods. It can be seen that the usage of the STVF suppression method improves the resolution of E^T_{miss} significantly. A detailed description of the measurement of the missing transverse energy and its performance can be found in reference [106].

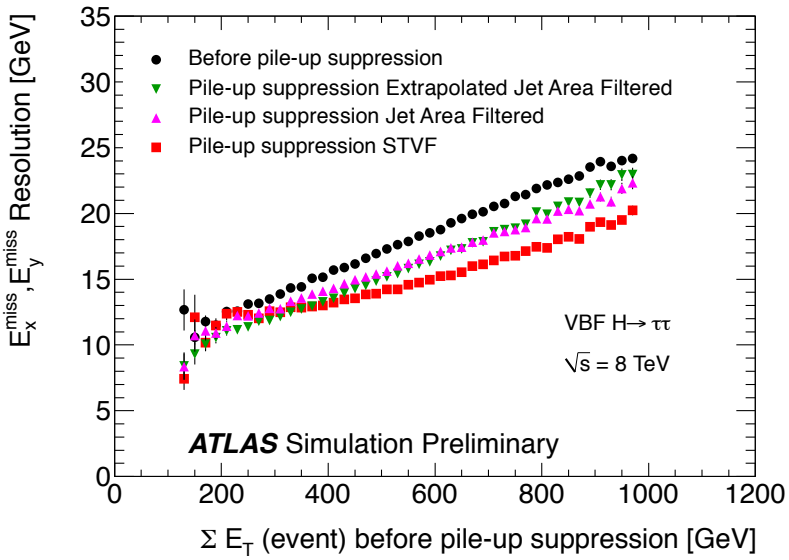


Figure 5.3: Resolution of the E^T_{miss} measurement for different pile-up suppression methods depending on the total transverse energy $\sum E^T$. The figure is taken from reference [106].

5.1.5 Overlap Removal Between Leptons and Jets

It is possible that one physics object is reconstructed twice meaning that it can happen that e.g a physical electron is reconstructed as an electron and a jet. Therefore it is necessary to perform an overlap removal to avoid this double counting. Two objects are considered to be overlapping if they point towards the same energy deposition in the calorimeter with a smaller separation than $\Delta R < 0.2$. The overlap removal of jets and leptons selects a object according to the following order of priority (highest to lowest): muons electrons and jet candidates.

5.2 Used Triggers

This analysis makes use of the same triggers as the SM $H \to \tau\tau$ analysis of ATLAS [37] in its fully leptonic channels.

Leptons in final state	Trigger	P^T requirement trigger	offline P^T requirements
1 e + 1 μ	EF_e12Tvh_medium1_mu8	$P_e^T > 12\,\text{GeV}$ $P_\mu^T > 8\,\text{GeV}$	$P_e^T > 15\,\text{GeV}$ $P_\mu^T > 10\,\text{GeV}$
	EF_e24vhi_medium1	$P_e^T > 24\,\text{GeV}$	$P_e^T > 26\,\text{GeV}$
	EF_e60_medium1	$P_e^T > 60\,\text{GeV}$	-
2 μ	EF_mu18_tight_mu8_EFFS	$P_{\mu_1}^T > 18\,\text{GeV}$ $P_{\mu_2}^T > 8\,\text{GeV}$	$P_{\mu_1}^T > 20\,\text{GeV}$ $P_{\mu_2}^T > 10\,\text{GeV}$
2 e	EF_2e12Tvh_loose1	$P_{e_1}^T > 12\,\text{GeV}$ $P_{e_2}^T > 12\,\text{GeV}$	$P_{e_1}^T > 15\,\text{GeV}$ $P_{e_2}^T > 15\,\text{GeV}$
	EF_e24vhi_medium1	$P_e^T > 24\,\text{GeV}$	$P_e^T > 26\,\text{GeV}$
	EF_e60_medium1	$P_e^T > 60\,\text{GeV}$	-

Table 5.1: List of used triggers, their P^T requirements and the offline P^T requirements.

The used triggers are shown in table 5.1 depending on the lepton content of the final state. All listed triggers in a given final state are combined with a logical *OR* meaning that an event is accepted if at least one of the triggers has fired. Additionally, a trigger matching procedure is applied ensuring that the event contains, after object reconstruction, a physics object corresponding to the one that caused the trigger to fire. Furthermore, the minimum P^T threshold of the leptons is increased such that they are larger than the minimal P^T requirements of the trigger, if the trigger was found to cause issues for the analysis due to turn-on effects. The raised (offline) P^T thresholds are listed in the last column of table 5.1.
Three different triggers are used for final states containing an electron and a muon in the final state: one di-lepton trigger asking for an electron of $P^T > 12\,\text{GeV}$ and a muon od $P^T > 8\,\text{GeV}$ and two single electron triggers asking for electrons with $P^T > 24\,\text{GeV}$ and $P^T > 60\,\text{GeV}$. This triggers with the higher P^T thresholds are introduced to compensate the efficiency loss of the trigger with a smaller P^T threshold.
It was found to be sufficient to use only one trigger for the final states containing two muons. This trigger is a di-muon trigger asking for two muons of $P^T > 18\,\text{GeV}$ and $P^T > 8\,\text{GeV}$.
Events containing two electrons in the final state are selected based on the decision of three different triggers. These triggers are one di-electron trigger asking for two electrons fulfilling $P^T > 12\,\text{GeV}$ and two single electron triggers asking for one electron with $P^T > 24\,\text{GeV}$ or $P^T > 60\,\text{GeV}$.

5.3 Basic Selection of the Vector Boson Fusion Topology

This chapter explains the cuts defining the basic VBF topology used in this analysis. All cuts shown in this section are implicitly applied in all following chapters. Table 5.2 lists all cuts and

	cut name	**cut value**	applied to data	applied to MC
Cut1	Good Runs List	flag	YES	NO
Cut2	Primary Vertex	$n_{\text{Tracks}} > 3$	YES	YES
Cut3	Jet Cleaning	flag	YES	YES
Cut4	Hot Cell Veto	flag	YES	NO
Cut5	LAR Error	flag	YES	NO
Cut6	No Tau	$n_\tau = 0$	YES	YES
Cut7	Dilepton	$n_{\text{lepton}} = 2$	YES	YES
Cut8	Trigger	flag	YES	YES
Cut9	Trigger Matching	depending on trigger	YES	YES
Cut10	Number Jets	$n_{\text{Jets}} > 1$	YES	YES
Cut11	P^T Jet 1	$P^T_{\text{Jet1}} > 40\,\text{GeV}$	YES	YES
Cut12	P^T Jet 2	$P^T_{\text{Jet2}} > 30\,\text{GeV}$	YES	YES
Cut13	$\Delta\eta$ Tagging Jets	$\Delta\eta_{\text{Jet1,Jet2}} > 2.2$	YES	YES

Table 5.2: List of cuts applied in the basic VBF selection. An event is kept if all criteria stated in the *cut value* column are met.

states whether they are applied on data and MC. The cuts can be classified into four different groups using the intended purpose of the cuts .
Cut1 to Cut5 form the first group including all cuts that are applied to ensure data quality. Cut1 checks if all subsystems of the ATLAS detector were working correctly during the events recording accepting only events whose luminosity block is part of the *Good Runs List* provided by the official *Data Preparation Group*. Events with four or more tracks pointing toward the primary vertex pass Cut2 suppressing events consisting mainly of pile-up. Cut3 rejects events containing *bad jets*. Jets are classified as *bad jets* if they are not originating from a real energy deposition in the calorimeters. *Bad jets* can be caused by hardware problems, cosmical rays or beam conditions. Cut4 removes events that contain at least one jet pointing toward a malfunctioning, so called *hot*, calorimeter cell, if the jet gained a significant amount of its energy from that cell. Additionally, events are rejected in Cut5, if an *LAr error* is detected during the recording of an event.
The second group of cuts is formed by Cut6 and Cut7. These cuts define the fully leptonic final state. Cut6 vetoes all events with a reconstructed hadronic τ lepton decay. This cut is used to ensure orthogonality of this analysis to the analyses considering hadronic decays of the τ lepton. Cut7 demands exactly two isolated leptons (here μ or e) of opposite charge in the final state.
Cut8 and Cut9 are both trigger related cuts. Cut8 ensures that one of the triggers used in the analysis has fired and therefore accepted the event. The triggers used in this analysis are discussed in section 5.2. The trigger matching of Cut9 ensures that the P^T of the object firing the trigger is sufficiently larger than the P^T threshold of the used trigger ensuring that the analysis is not affected by turn-on effects of the trigger. It ensures additionally that the event contains an object corresponding to the one that fired the trigger.
The VBF cuts are given by Cut10 to Cut13 selecting the VBF topology. These cuts exploit the signature of the signal produced in the VBF production mode in the detector, which is the presence of the tagging jets, as discussed in section 2.3.1. The first of the VBF cuts, Cut10, neglects therefore all events with less than two jets. This spares most of the signal produced in the VBF production

mode, as can be seen in figure 5.4 showing the jet multiplicity. The transverse momentum (P^T) spectra of the jets leading and sub-leading in P^T is shown in figure 5.5. The VBF signal tends to prefer larger values of P^T than most of the backgrounds. This motivates Cut11 and Cut12 demanding the transversal momentum of the leading and sub-leading jet to fulfill $P^T_{j1} > 40\,\text{GeV}$ and $P^T_{j2} > 30\,\text{GeV}$, respectively. The η separation $\Delta\eta_{jj}$ of the two jets leading in P^T is shown in figure 5.6. The given distributions indicate that the signal prefers large values of $\Delta\eta_{jj}$ for the VBF production mode, while all other shown processes prefer smaller values. Therefore, Cut13 is introduced demanding $\Delta\eta_{jj} > 2.2$. The same VBF cuts were used in the $H \to \tau\tau$ analysis of ATLAS [37] as a definition of the VBF topology.

	Cut9	Cut10	Cut11	Cut12	Cut13
$Z \to \mu\mu$	9361682 ± 5332.3	1318378 ± 1171.0	699710 ± 684.8	447354 ± 497.1	117391 ± 229.0
$Z \to ee$	6256001 ± 4280.6	796290 ± 901.8	476743 ± 549.7	282611 ± 406.1	75435 ± 182.9
Top quark	137513 ± 198.5	122904 ± 184.5	115842 ± 178.5	100500 ± 165.2	20279 ± 74.9
WW/WZ/ZZ	46753 ± 111.6	19711 ± 68.0	15418 ± 59.9	10380 ± 49.2	1967 ± 22.0
Fake	150986 ± 340.6	34225 ± 142.1	19232 ± 107.1	11147 ± 84.1	3539 ± 43.0
$Z \to \tau\tau$ (EMB)	178279 ± 113.5	25138 ± 45.6	15371 ± 37.0	9623 ± 29.7	2582 ± 15.1
VBF $H \to \tau\tau$	44 ± 0.1	35 ± 0.1	32 ± 0.1	25 ± 0.1	21 ± 0.1
ggH $H \to \tau\tau$	456 ± 1.4	122 ± 0.7	79 ± 0.6	51 ± 0.5	19 ± 0.3
VBF $H \to \tau\mu$	39 ± 0.2	30 ± 0.2	27 ± 0.2	21 ± 0.2	18 ± 0.2
ggH $H \to \tau\mu$	416 ± 2.6	121 ± 1.4	80 ± 1.1	48 ± 0.9	17 ± 0.5
VBF $H \to \tau e$	37 ± 0.2	28 ± 0.2	26 ± 0.2	20 ± 0.2	17 ± 0.1
ggH $H \to \tau e$	405 ± 2.5	115 ± 1.3	75 ± 1.1	44 ± 0.8	17 ± 0.5
Sum $H \to \tau\mu$	455 ± 2.6	151 ± 1.4	107 ± 1.2	69 ± 0.9	35 ± 0.6
Sum $H \to \tau e$	441 ± 2.5	143 ± 1.4	101 ± 1.1	64 ± 0.8	34 ± 0.5
Sum BG	16131714 ± 6851.1	2316802 ± 1498.4	1342428 ± 905.2	861690 ± 670.6	221233 ± 306.7
obs. data	16884410 ± 4302.3	2415566 ± 1554.2	1382108 ± 1175.6	858921 ± 926.8	236210 ± 486.0
Sig/BG	$\sim 3 \times 10^{-5}$	$\sim 6 \times 10^{-5}$	$\sim 8 \times 10^{-5}$	$\sim 8 \times 10^{-5}$	$\sim 1.5 \times 10^{-4}$

Table 5.3: Expected event yields and observed (obs.) data after enforcing the stated cuts and all cuts with a smaller number. The last line shows additionally the signal over background ration (Sig/BG). Please note that the dominant backgrounds $Z \to \mu\mu$, $Z \to ee$ and the top quark background are not normalized to data, hence the discrepancy between observed data and and the sum of the background (BG). The stated uncertainties correspond only to the statistical uncertainties.

The event yields after Cut9 to Cut13 are shown in figure 5.3. It can be seen that the VBF cuts reduce the expected background significantly. Only 1.4 % of the background events present after Cut9 pass Cut10 to Cut13. However, most of the signal ($\sim 95\,\%$) produced in the ggH production process and bout 55 % of the signal produced in the VBF production mode is lost as well. The overall signal to background ratio is nevertheless improved by a factor of roughly 5 without cutting on a variable that is sensitive to the decay of the Higgs boson.

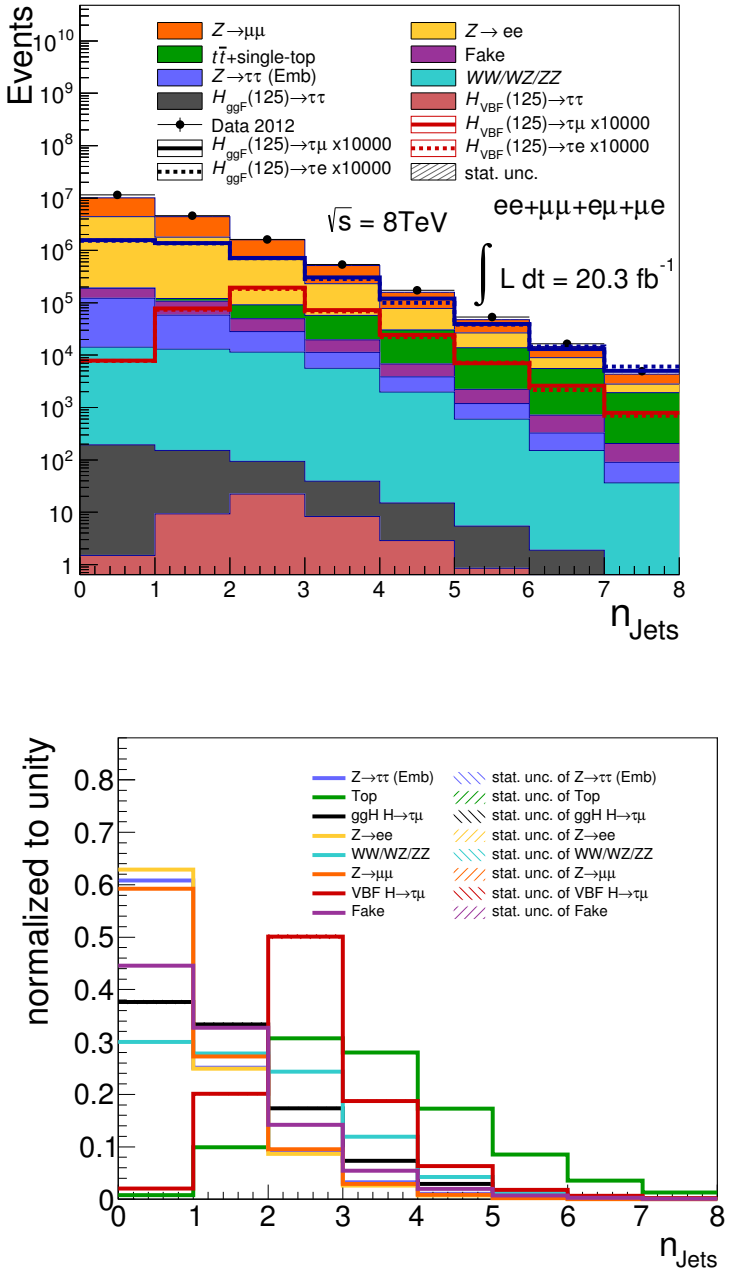


Figure 5.4: The distribution of the number of jets n_{Jets} is shown on the top after applying Cut9. The same distribution is shown normalized to unity in the bottom. The $H \to \tau e$ signal is not shown in the normalized plots since it has the same shape in this distribution as the $H \to \tau \mu$ signal.

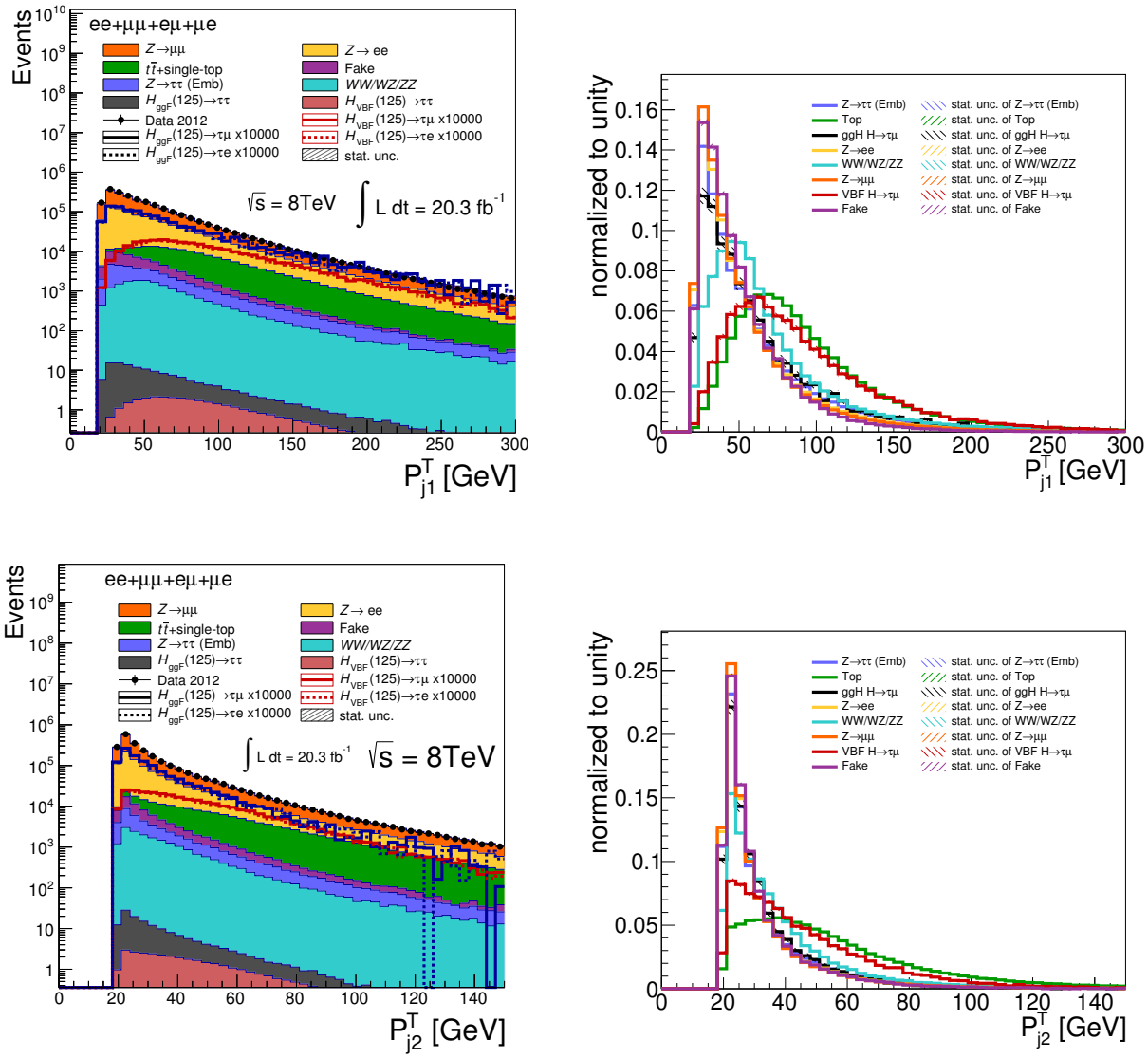


Figure 5.5: The left hand side shows distributions of the transverse momentum P^T of the jet leading (top) and sub-leading (bottom) in P^T after Cut10. The plots on the right hand side show the same distributions normalized to unity of some selected processes. The $H \to \tau e$ signal is not shown in the normalized plots since it has the same shape in the distribution as the $H \to \tau\mu$ signal.

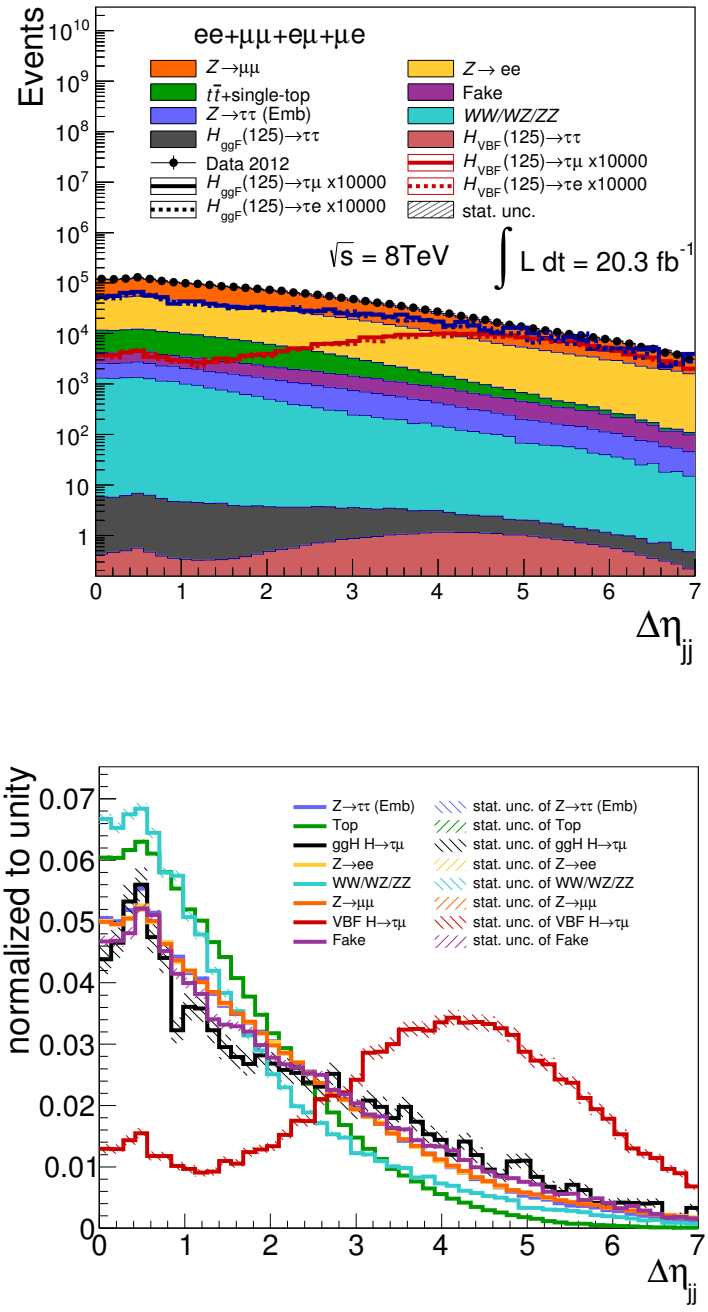


Figure 5.6: Distribution of the η separation $\Delta\eta_{jj}$ of the two jets leading in P^T after Cut10 on the top. The same distribution is shown normalized to unity in the bottom. The $H \to \tau e$ signal is not shown in the normalized plots since it has the same shape in this distribution as the $H \to \tau\mu$ signal.

Chapter 6

Lepton Labeling

This chapter describes the lepton labeling used in this analysis starting with the motivation. This is followed by a section discussing the underlying distributions and a section describing the procedure. Finally, the last section gives a description of the chosen lepton labeling as applied in the analysis. The lepton labeling is done after the basic VBF selection described in section 5.3. All shown distributions and presented results in this chapter are produced using the cuts of the basic VBF selection.

6.1 Motivation for Lepton Labeling

The LFV Higgs boson decays $H \to \tau\mu$ and $H \to \tau e$ considered in this thesis contain two charged leptons[1] in the fully leptonic final state. These charged leptons can be two electrons, two muons or an electron and a muon. The two charged leptons differ in their kinematic properties due to their different origins, as seen in section 2.3.1. One of the leptons, l_H, originates directly from the Higgs boson decay, while the other lepton, l_τ, originates from the decay of the τ lepton, as illustrated in figure 6.1. The leptons l_H and l_τ can both be electrons or muons depending on the decay channel

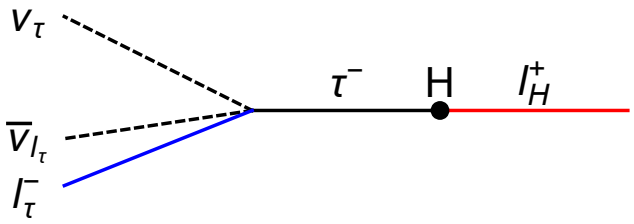


Figure 6.1: Systematical graph of an LFV Higgs boson decay illustrating the naming convention for the lepton originating from the Higgs boson decay l_H ($l_H = e, \mu$) and the lepton l_τ ($l_\tau = e, \mu$) originating from the decay of the τ lepton.

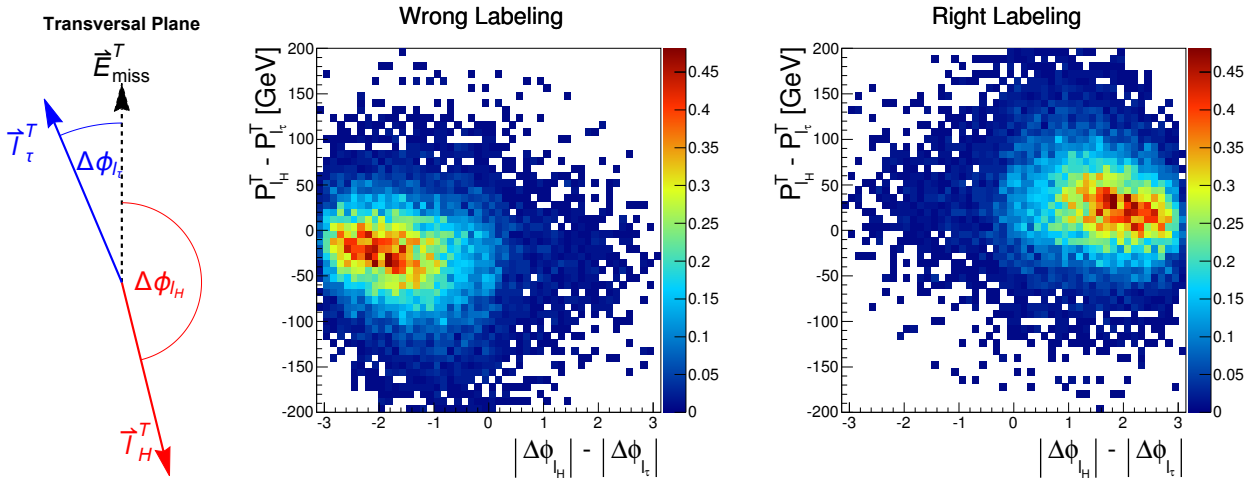
and final state.
Since l_τ and l_H differ significantly in their kinematic properties, it is important to identify them correctly in data. This identification of l_τ and l_H is called lepton labeling throughout this thesis to avoid confusion with the lepton identification in the event reconstruction.
Labeling the leptons properly is important for two reasons: First of all, it allows a better mass

[1] charged leptons refer here only electron and muon.

reconstruction since the angular correlation of l_τ and E_T^{miss}, caused by the energy carried away by the neutrinos, can be exploited. Furthermore, the kinematic differences between the leptons allow to separate the LFV signal from background processes like $H \to \tau\tau$ and $Z \to \tau\tau$, since these processes do not show on average such strong differences in their leptons' kinematics.

6.2 Underlying Distributions

The label of a lepton is neither depending on the lepton's flavor (muon or electron) nor on its charge, as it only gives information about the lepton's origin. Due to this, it is necessary to use



(a) (b) (c)

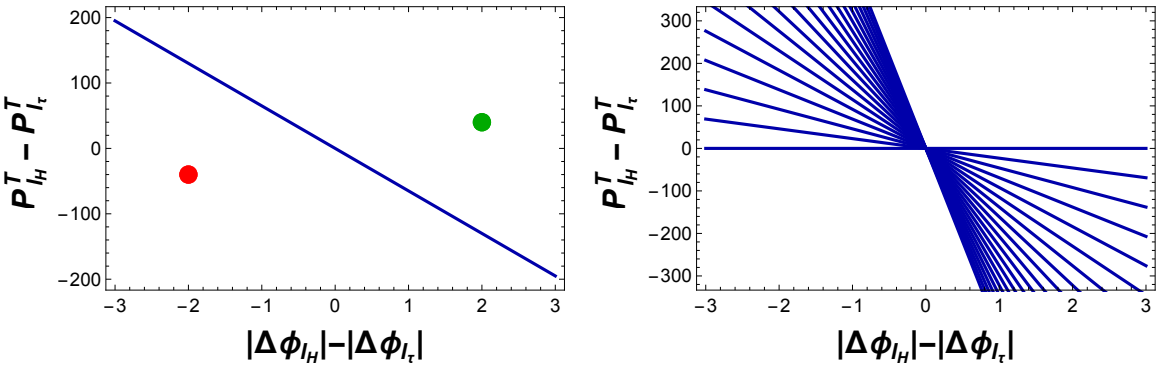
Figure 6.2: Figure (a) shows an illustration of the angles $\Delta\phi_{l_\tau}$ and $\Delta\phi_{l_H}$ between E^T_{miss} and the corresponding lepton in the transversal plane. Figure (b) and (c) show the distribution of the $H \to \tau\mu$ MC sample using truth information on the leptons origin for a wrong and right labeling in (b) and (c), respectively. The color code gives the percentage of events in each bin of the total amount of events.

variables in the labeling that are sensitive on the lepton's origin. Such variables are, for the signal studied in this analysis, the transverse momentum P^T of the leptons and the angle $\Delta\phi$ in the plane transverse to the beam axis between the lepton and $\vec{E}^T_{\text{miss}}$, as l_H receives on average a larger transverse momentum than l_τ, while $\Delta\phi_\tau$ tends to be smaller than $\Delta\phi_H$. The definition of $\Delta\phi_{l_\tau}$ and $\Delta\phi_{l_H}$ is illustrated in figure 6.2(a). Assuming the decay of a Higgs boson that does not have any transverse momentum itself and a perfect measurement of the neutrinos' transverse momentum by E^T_{miss} as well as a perfectly fulfilled collinear approximation would yield $\Delta\phi_\tau = 0$ and $\Delta\phi_H = \pi$. The absolute value of $P^T_{l_H}$ and $P^T_{l_\tau}$ is affected by a boost of the Higgs boson smearing the leptons P^T distributions. The measurement of $\Delta\phi_{l_\tau}$ and $\Delta\phi_{l_H}$ is affected by the poor resolution of the E^T_{miss} limiting their discriminating power. Therefore, the lepton labeling selected in this analysis makes only use of the difference of $\Delta\phi$ and P^T of the leptons, as a boost or a mismeasurement of E^T_{miss} affects both leptons in an event partially canceling their effects.
There are two possibilities to assign the labels l_τ and l_H to the two charged lepton is the final state. The right labeling gives the lepton originating from the Higgs boson decay the label l_H and the lepton originating from the decay of the τ lepton the label l_τ, while the wrong labeling swaps the

labels.
Figure 6.2(b) and 6.2(c) show the distribution of the $H \to \tau\mu$ MC sample in the $|\Delta\phi_{l_H}| - |\Delta\phi_{l_\tau}|$ vs. $P_{l_H}^T - P_{l_\tau}^T$ plane for wrong and right labeling, respectively, using truth information[2] on the leptons' origin. The lepton originating from the Higgs boson decay tends to have a larger P^T and $\Delta\phi$ than the lepton originating from the τ lepton decay. Due to this the distribution with the right labeling favors the upper right corner of the $|\Delta\phi_{l_H}| - |\Delta\phi_{l_\tau}|$ vs. $P_{l_H}^T - P_{l_\tau}^T$ plane, while the distribution with the wrong labeling prefers the lower left corner.

6.3 Lepton Labeling Procedure

The goal of the lepton labeling is to distribute the labels among the two leptons in the final state as correctly as possible, while ensuring that the procedure assigns the lepton labels in each event to avoid a loss of events due to the neglect of events that can not be labeled. The selected lepton



(a) (b)

Figure 6.3: Figure (a) and (b) give an illustration of the selection of the lepton label and the optimization of the lepton labeling, respectively. The blue line represents f_{Limit} at a tested value a, while the dots represent the two possible labelings. The green dot displays the selected labeling.

labeling of this analysis follows a simple procedure: a linear function f_{Limit} is defined as

$$f_{\text{Limit}}\left(|\Delta\phi_{l_H}| - |\Delta\phi_{l_\tau}|\right) = -a\left(|\Delta\phi_{l_H}| - |\Delta\phi_{l_\tau}|\right), \tag{6.1}$$

using a as a parametrization of the slope. The first possible labeling is tested by checking if the requirement of

$$f_{\text{Limit}}\left(|\Delta\phi_{l_H}| - |\Delta\phi_{l_\tau}|\right) > P_{l_H}^T - P_{l_\tau}^T\Big|_{\text{First possible labeling}} \tag{6.2}$$

is met by this labeling for a given a. This first possible labeling is selected as the final labeling, if the statement in equation (6.2) evaluates as true, else the second possible labeling is selected as the final labeling[3]. This is justified by the fact that distribution of the right labeling prefers the upper

[2] MC samples contain information on the origin of each particle of the event allowing to infer the true origin of the leptons. The term *truth information* refers to accessing these information.

[3] This procedure is technically speaking not well defined for all labels that fulfill the equation $f_{\text{Limit}}\left(|\Delta\phi_{l_H}| - |\Delta\phi_{l_\tau}|\right) = P_{l_H}^T - P_{l_\tau}^T$. In this case both labeling options lie exactly on top of f_{Limit} making it impossible to discriminate between the two labeling options using this procedure. The default selection for this case is the second tested labeling option, which corresponds to the assignment of a random label. However, this case is highly unlikely, as the numerical values used in the calculation have a high precision.

right corner in the $|\Delta\phi_{l_\tau}|$ vs. $P_{l_H}^T - P_{l_\tau}^T$ plane, as seen in figure 6.2(c). Figure 6.3(a) illustrates the procedure of the selection of the labeling. The two dots correspond to to labeling options and the blue line to f_{Limit}. The labeling lying above the blue line (green dot) is selected as the final label. This procedure is done for a range of possible values of a. The matching ratio r_m, that gives the fraction of the correctly labeled events from the total number of events, is calculated for each value of a and that with the best matching ratio is selected. This variation of a corresponds to a variation of the slope of f_{Limit} illustrated in 6.3(b).
Multiple functions for f_{Limit} have been tested and none of them performed better than the one presented in this section. Additionally, a Fisher discriminant was calculated using the the P^T and $\Delta\phi$ of the leptons individually resulting in the same matching percentage as the presented procedure. The presented procedure was selected over the Fisher discriminate due to its simplicity.

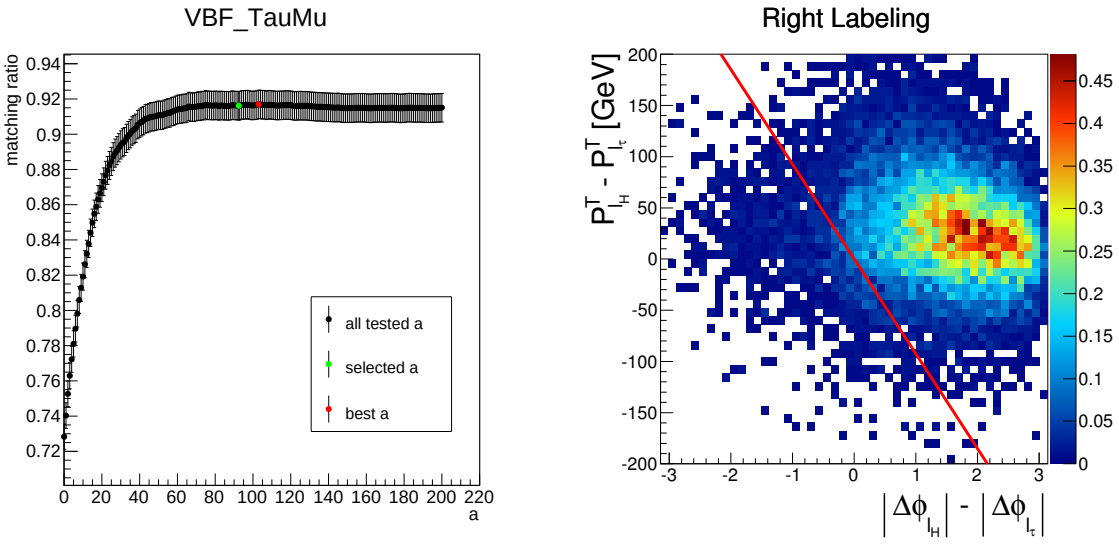
6.4 Selected Lepton Labeling

The lepton labeling was optimized for each considered production mode (ggH and VBF) and each signal ($H \to \tau\mu$ and $H \to \tau e$) separately. The parameter a is scanned in the interval of $[0, 200]$ with

MC sample	best a	r_m at best a	r_m at selected $a = 92.5$
VBF $H \to \tau\mu$	103	0.917 ± 0.008	0.916 ± 0.008
VBF $H \to \tau e$	82	0.913 ± 0.008	0.911 ± 0.008
ggH $H \to \tau\mu$	46	0.908 ± 0.028	0.901 ± 0.028
ggH $H \to \tau e$	53	0.879 ± 0.028	0.876 ± 0.028

Table 6.1: This table gives the best value of the parameter a and the matching ratio r_m for the best a and for the selected a. The given uncertainties on r_m are statistical nature.

an increment of 1 and in the range of $[200, 2000]$ with an increment of 5. It is not possible to have a different labeling for each production mode and each signal, since the labeling has to be applied for data as well, for which the production mode and signal process is not known. Therefore, one common value for a has to be selected. As this thesis focuses on the VBF topology, the mean of the best a-values in the VBF topology was selected, corresponding to $a = 92.5$. Table 6.1 gives the best a-value, the matching ratio r_m at the best a-value and the matching ratio at the selected a-value. It can be seen that the best matching ratio at the selected a-value is not far off the matching ratio at the best a-value. This is due to the fact that the matching ratio is distributed relatively flatly in the vicinity of the selected a, as can be seen exemplary in figure 6.4(a) showing the matching ratio depending on the parameter a for the $H \to \tau\mu$ sample in the VBF production mode. Figure 6.4(b) shows f_{Limit} using the selected value of $a = 92.5$ on top of the previously shown distribution of the right labeling using truth information.



(a) matching ratio depending on a for the $H \to \tau\mu$ sample in the VBF production mode

(b) illustration of f_{Limit} with the selected a for the $H \to \tau\mu$ sample in the VBF production mode

Figure 6.4: Figure (a) shows the matching ratio for a range of tested a-values. The red dot corresponds to matching ratio of the best a-value and the green dot to the matching ratio at the selected a. Figure (b) shows the f_{Limit} with the selected value of a represented by the red line on top of the leptons distribution with the right labeling using truth information.

Chapter 7

Final Event Selection Using BDTs

This chapter describes the final event selection. First an introduction on *Boosted Decision Trees* (BDT) [107–109] is given. This is followed by a section introducing the variables used in the analysis. The third section discusses the general analysis strategy that is used for the analysis of all final states. Finally, the last two sections present the analysis of the different and same flavor final states.

7.1 Boosted Decision Trees

BDTs are a widely used multivariate analysis tool in particle physics, as they are relatively easy to understand and therefore to control, and robust against both, statistical fluctuations and small changes in their settings. The presented analysis is based heavily on BDTs exploiting their ability to separate a small signal from a large background.
Like most multivariate analysis methods, a BDT evaluates how signal- or background-like an event appears to be, based on all information of a given set of variables. This distinguishes multivariate analysis methods from a traditionally cut-based analysis, where events are classified as signal- or background-like, based on the information of one variable at a time. Evaluating the information of the whole set of given variables simultaneously gives a multivariate analysis tool, like a BDT, the possibility to identify an event as signal even if it behaves very background-like in one variable, while it behaves signal-like in all other variables. Such an event would be classified as background and would consequently be rejected in a cut based analysis, since here an event is rejected if it was classified as background in one variable. Considering the whole set of variables simultaneously allows a BDT additionally to utilize correlations in the set of given variables. This is a big advantage of multivariate analysis over a cut-based analysis, especially when signal and background show similar distributions in a set of variables, but differ significantly in their correlations.
This section introduces briefly the concept of a BDT, followed by a discussion of the selected settings of the BDTs used in this analysis.

7.1.1 Introduction to Boosted Decision Trees

This analysis uses the implementation of BDTs in the *TMVA* framework [110]. A detailed report on the BDTs in the *TMVA* framework can be found in reference [110].

Decision Tree

A decision tree is a binary classifier that divides the parameter-space spanned by the set of input variables x_i into multiple regions and finally evaluates whether an event is signal- or background-like depending on the parameter-space region it is contained in. In the language of *TMVA* the parameter-space regions are called nodes.
Each decision tree has to be trained with a test sample and can afterwards be used for classification. During the training, the parameter-space is divided consecutively into subspaces using a training signal sample and background sample of same size[1]. All variables x_i are scanned for a threshold value x_i^t that divides the initial parameter-space in two sub spaces in such a way that one of the subspaces is as pure as possible with respect to the contribution of signal and background test events, according to the selected separation criterion for node-splitting. Eventually, the $x_i^{t,\text{best}}$ is selected from the set of all x_i^t that creates, after splitting the initial parameter at $x_i^{t,\text{best}}$, the purest subspace. This $x_i^{t,\text{best}}$ is used for splitting the initial parameter-space for the first time. The procedure is repeated on the newly formed subspaces until a stopping requirement is met. Such stopping criteria can be for example the amount of events that are left in a parameter-space region, or the number of consecutive splits that have been performed, the so-called depth of the decision tree.
After the splitting procedure is completed, each parameter-space region is classified as signal or background depending on the content of the training samples in the region. A region is classified as signal if the weighted sum of all events originating from the signal training sample is larger than the weighted sum of the events originating from background training sample in that specific region. The training of the decision trees is finished after the classification of the parameter-space region. An event is classified simply by evaluating in which parameter-space region it is contained. Due to this, a decision tree classifies an event discretely as either signal or background, but is not able to give feedback on the accuracy of this classification in terms of how signal- or background-like an event is. A graphical representation of a decision tree is given in Figure 7.1.

Boosting of Decision Trees

Boosting expands the idea of a single decision trees to extract more information than just a discrete classification of events from the given set of variables. Multiple decision trees are trained with the same training sample reweighed for each individual tree. A weighted average is used to extract a common response of the BTD from the classifications of the individual decision trees. This common response, the so called BDT score, is mapped to the interval of -1 to 1 and enables therefore BDTs to give a continuous evaluation on how signal- or background-like an event is. Events that are very signal-like will receive a BDT score close to 1, while very background-like events will most likely receive a score close to -1.
BDTs are more often used than decision trees, since they usually outperform a single decision tree significantly. Boosting stabilizes additionally the classifiers output with respect to statistical fluctuations in the training sample.

7.1.2 Settings of Boosted Decision Trees Used in this Analysis

The BDTs of the *TMVA* framework have adjustable settings allowing to customize them for an optimal performance on the given input. An optimal setting of a BTD has to meet two criteria. On

[1] The actual amount of events in the two samples do not have to be equal, as long as their sum of weights is equal.

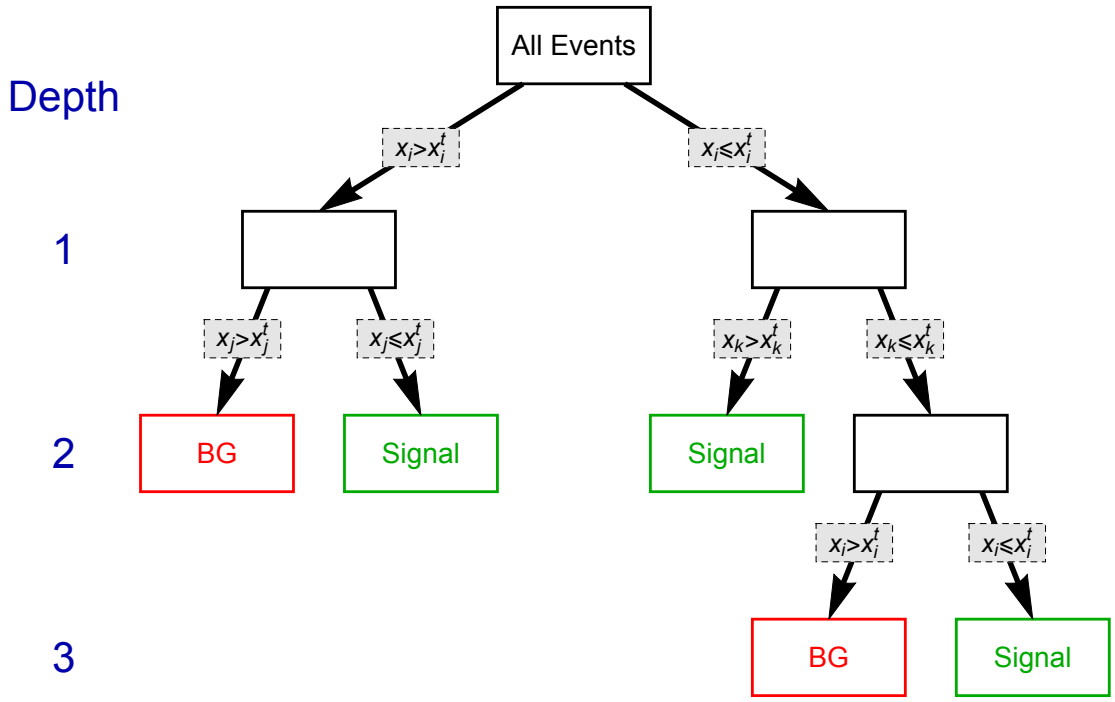


Figure 7.1: Graphical representation of a decision tree. The solid boxes correspond to regions of the parameter-space, or so called nodes. The black solid boxes are internal nodes, that are split further, as none of the stopping criteria has been met. The stopping criteria do not allow to split final nodes, shown in red and green. They are classified as signal or background depending on their training sample's content. The blue numbers give the depth of the decision tree, that is defined as the count of consecutive splitting that had to be performed to reach that node.

one hand, it has to provide an optimal separation of signal and background in the BDT score, while it has to ensure on the other hand that the BDT is not over-trained. An over-trained BDT becomes sensitive to sample-specific properties of the training samples that are not physical features of the signal or background processes considered in the training, like statistical fluctuations of the training sample. This decreases of course the performance of the BDT on other samples and therefore has to be avoided.
The parameters of the BDT were optimized for the μe final state but used in all final states, as they were found to perform well. Furthermore, the BDTs used in this analysis were found to perform very stably under a variation of the setting parameters in a reasonable range.
The meaning of the settings and their optimized values are listed in the following.

- **Number of trees = 300**. This setting allows to select the number of decision trees that are used in the BDT. An increased number of decision trees increases in general the performance of a BDT with respect to the separation of signal and background, but it makes the BDT vulnerable to over-training, if it is too large.

- **Minimal node size = 1 %**. This parameter describes the minimal amount of events that have to be contained by a parameter-space region in terms of the percentage of the training sample size. A parameter-space region is not split if the resulting region contains less events.

A small node size enhances the BDT's ability to discriminate between signal and background, but it makes it vulnerable to over-training.

- **Maximal depth = 5**. This parameter stops the growth of a BDT, if its depth reaches the maximal depth. A large maximal depth allows the BDT to discriminate better between signal and background, but can cause over-training.
- **Number of cuts = 20**. The variable range is scanned equidistantly for each variable x_i during the selection of the best value x_i^t. This setting gives the number of cuts that are tested during this process for each variable. A large number of tested cuts can enhance the performance of the BDT, but increases vulnerability towards over-training.
- **Boost type = Grad**. There are multiple boosting algorithms to select from in the *TMVA* package. The Grad boosting algorithm was selected since it was found perform well in the SM $H \to \tau\tau$ analysis [37] of the ATLAS collaboration.
- **Shrinkage = 0.05**. This parameter describes the learning rate of the Grad boosting algorithms.
- **Separation criterion for node-splitting = GiniIndex**. The Gini index [111] G is a measure for inequality. It is defined as $G = p(1-p)$ with $p = \frac{n_S}{(n_S+n_B)}$ where the number of signal and background events are denoted by n_S and n_B, respectively.

The *TMVA* package allows to adjust more parameters than the listed ones. All parameters that are not listed were used in their default settings.

7.1.3 Trainings Strategy for Boosted Decision Trees

The BDTs used in this analysis were trained with weighted events. These weights are calculated such that the sum of weights corresponds to the expected event yield of each contribution at the stage of training. This ensures that the BDT is trained focusing on the important backgrounds of the analysis.
For the training of the BDT it is necessary to introduce two classes of training events, signal and background. The sum of all weighted background events was used as the background class in training, while only the signal MC of the VBF production mode was used as the signal class, neglecting the signal MC of the ggH production mode. This allows the BDT to focus on the properties of the VBF production mode without suppressing actively the signal of the ggH production mode.

Cross Training

The training and evaluation of a BDT should not be done on the same set of events to avoid that a BDT is gaining its discriminating power from using unphysical, dataset specific properties, like statistical fluctuations, rather than from real physical properties of signal and background. If this is not considered, a sufficiently over-trained BDT can gain an unphysical arbitrarily strong separation power between signal and background that is specific to the dataset used in the training.
This analysis splits therefore all events into two equal sets, s_1 and s_2. These sets s_1 and s_2 are then used to train two BDTs, BDT A and BDT B, where BDT A is trained using only events of set s_1 and BDT B is trained using events contained in set s_2. BDT A is now evaluated on set s_2 and BDT B on set s_1. This procedure is called cross training and allows to use the all the events in the training and evaluation of the BDTs without doing the evaluation and training of the BDT

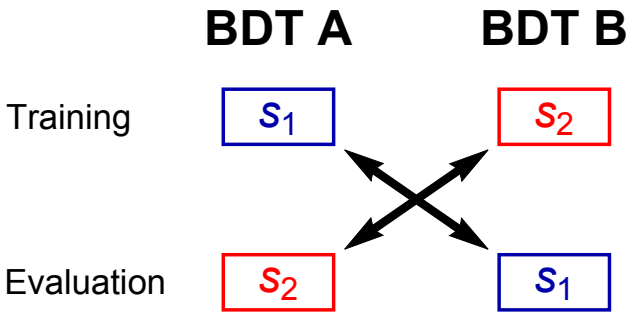


Figure 7.2: Illustration of cross training.

on the same set of events. Being able to use all events in the evaluation of the BDTs is important for this analysis, as the available amount of MC simulation for signal and background processes are very limited in the phase space corresponding to the VBF topology and reducing the amount of available MC events by 50% would affect the analysis.
BDTs can be checked for effects of over-training by evaluating the BDT on both the sample used in training and the test sample used in the evaluation of the BDT. An over-trained BDT is sensitive to the specific non-physical properties of the training sample and will therefore perform significantly better on the training sample than on the test sample. This means that the distribution of the BDT score of training and test sample will differ significantly for an over-trained BDT. This analysis accepts therefore only BDTs that show a good agreement between the distribution of the BDT score of the training and test sample. A visualization of the cross training is given in figure 7.2.

Variable Selection

The main goal of the variable selection is keeping the BDT as simple as possible in terms of the amount of considered input variables while maintaining maximal performance. *TMVA* produces a ranking of the input variables based on the variable's importance. The importance of a variable is a measure for how much the BDT makes use of a certain variable combining information on the separation power and number of usages of a variable.
The variable selection of the BDT is started by training a BDT that uses all available variables. Then the lowest ranked variable is neglected and a new BDT is trained with the new set of variables. The performance of the new BDT is compared to the performance of the initial BDT that makes use of all input variables. If both BDTs perform similarly well, the lowest ranked variable is neglected and the procedure is repeated with the variable that is ranked lowest in the ranking of the new BDT. Eventually, only variables that are not neglected during this procedure are used in the training of the final BDT. This final BDT is then used in the analysis.

Constraining the Variable Range

A BDT of the *TMVA* package scans the whole range of a variable spanned by the highest and lowest value of the variable equidistantly with the specified number of cuts to find the optimal value for splitting the initial parameter space into two subspaces. This procedure is problematic for variables that have extreme outliers in their distributions, as these extreme outliers increase the scanned variable range drastically. This yields that the distance between the tested scan point in that variable can get bigger than the characteristic features of the distribution of the variable, making the BDT insensitive to them. This is especially problematic for variables like m_{coll} and

$M_{\text{coll},\tau\tau}$ that are introduced in section 7.2, since they are defined as a fraction which means that a sufficient small denominator can yield arbitrary large values of this variables.
This analysis deals with this issue by setting the outliers to a fixed value in the training of the BDT. Two threshold value v_{min} and v_{max} are calculated for each variable v, such that 2.5 % of all events have a larger (smaller) value of v than v_{max} (v_{min}). With that in hand, all events with $v < v_{\text{min}}$ ($v > v_{\text{max}}$) are set to v_{min} (v_{max}) in the training of the BDT. This allows to make the BDT sensitive to variables with extreme outliers, while using all events in the training.

7.2 Variables Used in Selection

This section introduces all variables which are used in the analysis presented in this thesis. Multiple other variables were tested, but found to contribute insignificantly and were therefore not considered.

7.2.1 Missing Transverse Energy

Two different measurements for the missing transverse energy are used in this analysis.

Missing Transverse Energy E^T_{miss}

The standard transverse energy denoted as E^T_{miss} whose measurement was described in section 5.1.4.

Missing Transverse Energy of High P^T Objects E^T_{miss}(hpto)

A missing transverse energy calculated using only objects with a high transverse momentum, namely all jets and leptons (= e,μ) above the P^T threshold for these object described in section 5.1. It is defined as

$$\vec{E}^T_{\text{miss}}(\text{hpto}) = -\sum_{l=\text{leptons}} \vec{P}^T_l - \sum_{j=\text{jets}} \vec{P}^T_j, \tag{7.1}$$

$$E^T_{\text{miss}}(\text{hpto}) = \left|\vec{E}^T_{\text{miss}}(\text{hpto})\right|. \tag{7.2}$$

This version of the missing transverse energy is less affected by pile-up than the regular version of the missing transverse energy, as it is only calculated from the high P^T objects.

7.2.2 Variables Describing Kinematics of the Leptons

The leptons l_H and l_τ are the only directly measured particles originating from the LFV decay of the Higgs boson considered in this analysis. They contain therefore most of the available information on the Higgs boson decay. Due to this, variables describing the kinematics of the leptons are used widely in this analysis.

Transverse Momentum $P^T_{l_H}$ of l_H

The transverse momentum of l_H, denoted by $P^T_{l_H}$, tends to be larger in signal processes than in background processes like $H \to \tau\tau$ and $Z \to \tau\tau$.

Invariant Mass of the Lepton System m_{ll}

The invariant mass of the lepton system formed by l_H and l_τ is a very helpful variable discriminating against $Z \to ee$ and $Z \to \mu\mu$, as m_{ll} corresponds, for these processes, to the reconstructed mass of the Z boson, causing a narrow resonance in m_{ll} at the Z bosons mass. Additionally it is an input variable for the reconstruction of the Higgs boson mass that is described in section 7.2.4.

Transverse Mass of the Lepton System m_{ll}^T

The variable m_{ll}^T denotes the transverse mass of the lepton system formed by l_H and l_τ. The signal favors larger values of this variable than most background processes.

Separation ΔR of Leptons ΔR_{ll}

The separation of l_H and l_τ in R is given by

$$\Delta R_{ll} = \sqrt{(\eta_{l_H} - \eta_{l_\tau})^2 + (\phi_{l_H} - \phi_{l_\tau})^2}. \tag{7.3}$$

Impact Parameter $d0_{l_\tau}$ and its Significance $S(d0_{l_\tau})$

While τ leptons decay too fast to be detected directly, their existence can still leave a small footprint in the detector. This footprint can be measured by using the impact parameter $d0_{l_\tau}$ describing the

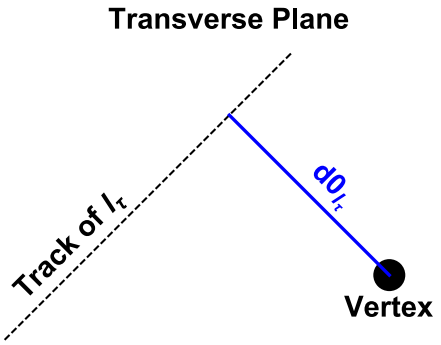


Figure 7.3: Illustration of the variable $d0_{l_\tau}$.

minimal distance between the track of l_τ and the reconstructed primary vertex of the event in the transverse plane visualized in figure 7.3. A strong enough boost of the τ lepton can displace the track of the lepton originating from the τ decay from the reconstructed vertex, as such a τ lepton can travel a small, but measurable distance in the detector before it decays, causing the impact parameter $d0$ not to equal zero. The signal tends to favor larger values for $d0_{l_\tau}$, since here l_τ is very likely originating from a τ decay, while the lepton labeled as l_τ is not originating from a τ decay for background processes like $Z \to \mu\mu$ and is therefore favoring small values of $d0_{l_\tau}$. The benefit of this variable is limited, as it is not perfectly modeled in MC and the resolution of the ATLAS detector is at the same order of magnitude as the measured effect. It provides nevertheless a good discrimination of the signal from the $Z \to \mu\mu$ and $Z \to ee$ background.
The significance $S(d0_{l_\tau})$ is defined by

$$S(d0_{l_\tau}) = \frac{|d0_{l_\tau}|}{\sigma(d0_{l_\tau})}, \tag{7.4}$$

using the uncertainty $\sigma\left(d0_{l_\tau}\right)$ of $d0_{l_\tau}$. This variable takes the limited resolution of $d0_{l_\tau}$ into account. Both variables are used in the analysis.

Difference in ϕ of l_τ and E^T_{miss} Denoted as $\Delta\phi(l_\tau, E^T_{\text{miss}})$

$\Delta\phi(l_\tau, E^T_{\text{miss}})$ is defined as

$$\Delta\phi(l_\tau, E^T_{\text{miss}}) = \left|\phi_{l_\tau} - \phi_{E^T_{\text{miss}}}\right|. \tag{7.5}$$

LFV Higgs decays favor a small value for $\Delta\phi(l_\tau, E^T_{\text{miss}})$ since E^T_{miss} is mainly caused by the neutrinos originating from the τ decay, that are almost collinear to l_τ due to the strong boost of the τ lepton in these decays. All other background processes do not show such a strong correlation between E^T_{miss} and l_τ causing them to have a wider distribution of $\Delta\phi(l_\tau, E^T_{\text{miss}})$ than the signal.

Transverse Mass of E^T_{miss} and l_τ Denoted as $m^T(E^T_{\text{miss}}, l_\tau)$

The transverse mass of the system l_H and E^T_{miss}, denoted by $m^T(E^T_{\text{miss}}, l_\tau)$, is a reconstruction of the transverse τ mass for the signal processes and favors therefore small values of $m^T(E^T_{\text{miss}}, l_\tau)$. The $m^T(E^T_{\text{miss}}, l_\tau)$ variable does in general not correspond to a reconstruction of the transverse mass of a τ lepton for background processes. This is in particular true for the top quark background for which $m^T(E^T_{\text{miss}}, l_\tau)$ tends to be large. This variable is therefore a powerful variable for the different flavor final states that are dominated by the top quark background.

7.2.3 Variables Describing the Kinematics of the Jets

This analysis uses only variables describing the two leading jets in P^T. These jets correspond with a high probability to the tagging jets in the VBF signal samples and are therefore sensitive to the production of the Higgs boson in the VBF production mode. Due to this, the variables describing the jet kinematics provide a possibility to discriminate between signal and background that is independent of the decay of the Higgs boson.

η Separation of Tagging Jets $\Delta\eta_{jj}$

The separation in η of two leading jests j_1 and j_2 is given by

$$\Delta\eta_{jj} = \left|\eta_{j1} - \eta_{j2}\right|. \tag{7.6}$$

This variable was already used in the basic VBF cuts of this analysis described in section 5.3 defining the VBF topology. Despite that, it is still a powerful variable, as the signal in the VBF production mode favors large values of $\Delta\eta_{jj}$, while most backgrounds prefer smaller values, even after the basic VBF cuts.

Invariant Mass of the Tagging Jet System m_{jj}

The invariant mass of the tagging jet system m_{jj} formed by the the two jets leading in P^T is another variable sensitive to the VBF production mode, as the signal produced in the VBF production mode favors large values for m_{jj} while most other backgrounds favor small values.

7.2.4 Variables from the Higgs Boson Mass Reconstruction

The reconstructed mass of the Higgs boson is one of the most sensitive variables used in this analysis. The so-called collinear approximation is used to reconstruct the Higgs boson mass described first in reference [112]. Two different approaches are introduced and used in this analysis. The first one is a collinear mass m_{coll} tailored for the kinematics of the LFV Higgs decay assuming a decay of the form $H \to \tau l_H$, while the second one is the collinear mass $m_{\text{coll},\tau\tau}$ of the $H \to \tau\tau$ analysis of ATLAS assuming a decay of the Higgs boson into two τ leptons.
Both methods reconstruct the mass of a Higgs boson in LFV decay correctly, while they give in general different results for processes containing two τ leptons such as the SM decays $H \to \tau\tau$ and $Z \to \tau\tau$. This can be seen in figure 7.4 showing the mass distribution of both mass reconstructions for these processes. These plots show furthermore that $M_{\text{coll},\tau\tau}$ is not able to separate between the LFV and SM Higgs decay, while m_{coll} separates them. Therefore both collinear masses are used in

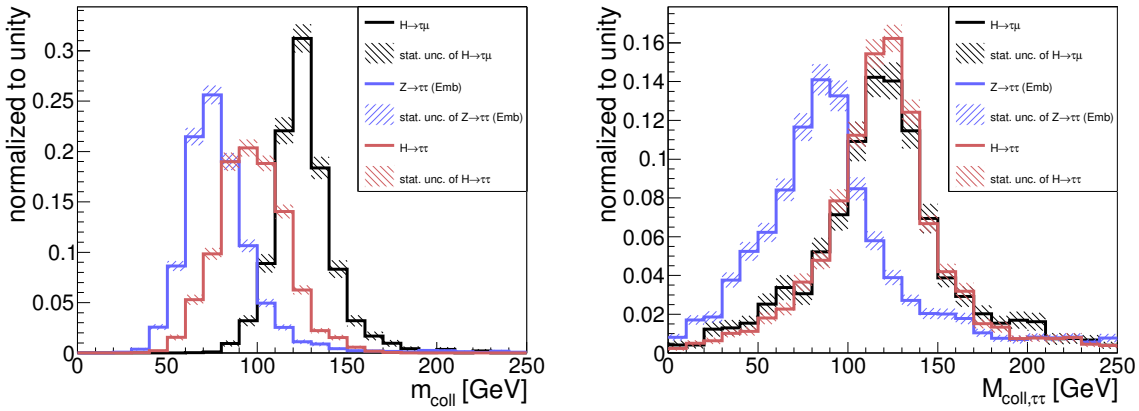


Figure 7.4: Comparison of m_{coll} (left) and $M_{\text{coll},\tau\tau}$ (right) for the LFV signal $H \to \tau\mu$, the decay $H \to \tau\tau$ of the SM Higgs boson and the $Z \to \tau\tau$ background after the event selection of the μe final state described in section 7.4.1. All shown distributions are normalized to unity and the stated errors correspond to the statistical uncertainties.

the analysis, giving the BDTs the option to exploit the different correlation between the two mass variables for the LFV signal and the SM background.
The crucial part in the calculation of m_{coll} ($m_{\text{coll},\tau\tau}$) is the reconstruction of the initial momentum (momenta) of the τ lepton(s). Once the momentum (momenta) of the τ lepton(s) is reconstructed it is easy to infer the mass of the Higgs boson by calculating the invariant mass of the system formed by the reconstructed τ and l_H (by both reconstructed τ leptons). Both use the assumption that all decay products of the τ leptons are collinear to the original τ lepton, hence the names collinear approximation and collinear mass.

Collinear Mass m_{coll} and x Assuming LFV Higgs Decays

This version of the collinear mass, that was taken from reference [35][2], assumes that the decay of the Higgs boson contains exactly one τ lepton that is boosted significantly, such that the trajectory

[2]The collinear mass used in the analysis of reference [35] uses the same assumptions and derives the same final formula for the collinear mass, but its intermediate steps differ from the derivation presented in this thesis.

of its decay products are collinear to the trajectory of the original τ lepton and that the neutrinos originating from the τ decay are the only contribution to E^T_{miss}. Within these assumptions, $\vec{E}^T_{\text{miss}}$ has to be parallel to l_τ in the transverse plane meaning that a contribution to $\vec{E}^T_{\text{miss}}$ that is not parallel to l_τ is considered to be a non-physical contribution.
Therefore the transverse momentum $P^T_{\tau,\text{reco}}$ of the τ lepton is reconstructed by

$$P^T_{\tau,\text{reco}} = P^T_{l_\tau} + \Pi\left(\vec{E}^T_{\text{miss}}\right), \tag{7.7}$$

where the projection Π of $\vec{E}^T_{\text{miss}}$ on $\vec{P}^T_{l_\tau}$ was introduced. This projection removes the contributions

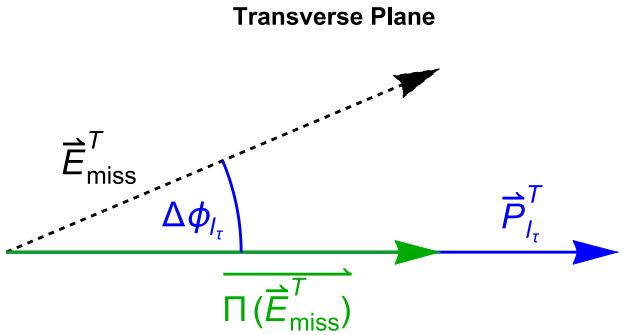


Figure 7.5: Illustration of the the projection of $\vec{E}^T_{\text{miss}}$.

of $\vec{E}^T_{\text{miss}}$ that are not parallel to $\vec{P}^T_{l_\tau}$. Figure 7.5 illustrates the projection of $\vec{E}^T_{\text{miss}}$.
With this in hand it is possible to define the the ratio of the τ lepton's transverse momentum carried by l_τ as

$$x = \frac{P^T_{l_\tau}}{P^T_{\tau,\text{reco}}}. \tag{7.8}$$

This yields

$$\vec{P}_{\tau,\text{reco}} = \frac{\vec{P}_{l_\tau}}{x}, \tag{7.9}$$

making use of the collinear assumption. With this in hand, the momentum of the initial τ lepton is fully reconstructed and the Higgs boson mass is reconstructed as

$$m_{\text{coll}} = \frac{m_{ll}}{\sqrt{x}}, \tag{7.10}$$

using the additional assumption that the masses of the τ lepton, of l_H and of l_τ are small compared to their momenta and can therefore be neglected.
The collinear mass m_{coll} is a powerful variable in all final states and is used in all BDTs. The variable x has some discrimination power itself and is therefore used as well.

Collinear Mass $M_{\text{coll},\tau\tau}$ for the SM $H \to \tau\tau$ Decay, x_{l_H} and x_{l_τ}

This version of the collinear mass is used to reconstruct the mass of the Higgs boson in the $H \to \tau\tau$ decay. It assumes as well that all decay products of a τ lepton are collinear with the initial τ lepton and that $\vec{E}^T_{\text{miss}}$ is solely caused by the neutrinos originating from the τ decay. For this decay of the Higgs boson it is no longer possible to assume that $\vec{E}^T_{\text{miss}}$ is parallel to one of the leptons, since

$\vec{E}^T_{\text{miss}}$ corresponds to the vectorial sum of the momenta of all the neutrinos from the decay of both τ leptons. To take this into account x_{l_H} and x_{l_τ} are introduced by

$$x_{l_H} = \frac{P^T_{l_H}}{P^T_{\tau_1}}, \qquad x_{l_\tau} = \frac{P^T_{l_\tau}}{P^T_{\tau_2}}, \tag{7.11}$$

describing the fraction of the initial τ leptons transverse momentum carried by the visible decay products l_H and l_τ. Using the collinear approximation to reconstruct the initial transverse τ momenta and conservation of momentum yields

$$\vec{P}^T_{\tau_1} + \vec{P}^T_{\tau_2} = \frac{\vec{P}^T_{l_H}}{x_{l_H}} + \frac{\vec{P}^T_{l_\tau}}{x_{l_\tau}} + \vec{E}^T_{\text{miss}}, \tag{7.12}$$

assuming that $\vec{E}^T_{\text{miss}}$ is caused solely by the neutrinos from the decay of the τ leptons. This equation is obviously valid for the x and y component of the transverse plane providing two independent equations that allow to solve for x_{l_H} and x_{l_τ} giving the result

$$x_{l_H} = \frac{P^x_{l_H} P^y_{l_\tau} - P^y_{l_H} P^x_{l_\tau}}{P^x_{l_H} P^y_{l_\tau} - P^y_{l_H} P^x_{l_\tau} + P^y_{l_\tau} E^x_{\text{miss}} - P^x_{l_\tau} E^y_{\text{miss}}}, \tag{7.13}$$

$$x_{l_\tau} = \frac{P^x_{l_\tau} P^y_{l_H} - P^y_{l_\tau} P^x_{l_H}}{P^x_{l_\tau} P^y_{l_H} - P^y_{l_\tau} P^x_{l_H} + P^y_{l_H} E^x_{\text{miss}} - P^x_{l_H} E^y_{\text{miss}}}. \tag{7.14}$$

By neglecting the masses of the τ leptons, l_H and l_τ and using once more the collinear approximation, it is possible to reconstruct the mass to

$$M_{\text{coll},\tau\tau} = \frac{m_{ll}}{\sqrt{x_{l_H} x_{l_\tau}}}. \tag{7.15}$$

This expression is only well defined if $x_{l_H} x_{l_\tau} > 0$. This issue is solved in this analysis by omitting the negative sign of $x_{l_H} x_{l_\tau}$ by calculating $M_{\text{coll},\tau\tau}$ with $|x_{l_H} x_{l_\tau}|$ and setting finally $M_{\text{coll},\tau\tau}$ to the negative of the previously calculated value. Due to this, it is possible to calculate $M_{\text{coll},\tau\tau}$ for all events at the price of introducing negative values for $M_{\text{coll},\tau\tau}$[3].
The analysis makes use of both $M_{\text{coll},\tau\tau}$ and x_{l_H}. The LFV decays $H \to \tau\mu$ and $H \to \tau e$ tend to peak in the vicinity of 1 in x_{l_H} since there is no corresponding τ lepton for l_H in these decays meaning that all momentum of this decay branch is visible, corresponding to $x_{l_H} = 1$, while processes like $H \to \tau\tau$ and $Z \to \tau\tau$ have neutrinos on both branches of the decay causing x_{l_H} to be in general smaller than 1.

7.2.5 B-Tagging

The special topology of jets originating from b quarks, so called b-jets, allows to discriminate them from jets originating from other quarks. This identification of b-jets is called *b-tagging* and is done in this analysis using the tagging algorithm *MV1* [113]. This tagging algorithm calculates for each jet, with $P^T > 25\,\text{GeV}$ and $|\eta| < 2.5$, a score between 0 and 1 with a high score corresponding to a high likelihood for the jet being a b-jet. A jet is tagged as a b-jet in this analysis, if its score is above the threshold of 0.3511 corresponding to a working point with a b-tagging efficiency of 80% for true b-jets. This working point is stricter than the working point at 70% of the SM $H \to \tau\tau$ analysis allowing eventually a better suppression of the top quark background that is the dominant background in the same flavor final states.

[3] This negative sign in $M_{\text{coll},\tau\tau}$ has no physical meaning. It is just a technical solution to cure $M_{\text{coll},\tau\tau}$ in the ill defined parameter space of $x_{l_H} x_{l_\tau} < 0$.

bTag

The discrete variable $bTag$ is a global variable for the events. It is 0 if no jet in the event is tagged as a b-jet and 1 if one or more jets in the event are tagged as b-jets. The signal process does not contain any b-jet at tree level. So, signal events will most likely have $bTag = 0$, while the top quark background contains b-jets at tree level, as the top quark decays almost exclusively to a b quark and a $W^{\pm}$ boson. The variable $bTag$ will therefore most likely be equal to 1 for the top quark background. Due to this, the variable $bTag$ is very powerful in discriminating the top quark background from the signal and therefore a crucial ingredient to the different flavor final states, that are dominated by the top quark background.

7.3 Analysis Strategy

This section discusses and motivates the general strategy selected in the presented analysis. First, a motivation for the usage of BDTs is given, followed by an explanation of the handling of BDTs throughout this analysis. Finally, the four final states are introduced and discussed.

7.3.1 Use of Boosted Decision Trees

The biggest challenge in this analysis is dealing with the tiny signal to background ratio and the potentially very small signal yield. The first step taken in this analysis to deal with this issue was to focus completely on the VBF topology exploiting the clear signature of the VBF production mode of the Higgs boson to enhance the signal to background ratio.
Nevertheless, the signal to background ratio is with approximately 2.2×10^5 expected background and 34 (35) expected[4] $H \to \tau\mu$ ($H \to \tau e$) signal events still poor after the VBF cuts of the basic selection described in section 5.3. This small signal to background ratio and the overall small signal yield are a challenging combination for a traditional cut-based analysis. A cut-based analysis makes it in this scenario on one hand necessary to cut harshly on discriminating variables to gain a suitable signal to background ratio and create therefore a sensitive analysis, while it is on the other hand necessary to spare the precious signal by selecting soft cuts to avoid loosing sensitivity due to the small remaining signal yield. It is obvious that both are hard to accomplish simultaneously and therefore its is usually necessary to compromise between both.
This analysis makes use of the multivariate analysis tool of BDTs to avoid this issue. A BDT-based strategy provides for this analysis two main advantages over a cut-based strategy that are explained in the following.
The first advantage of a BDT is that no signal or background events are lost due to cuts, while signal and background are separated in the BDT score. Therefore, the BDT score has two key features: the high BDT score region is enriched in signal enhancing the significance of the analysis, while the lower region of the BDT is almost signal free and contains most of the background events allowing to infer constraints on the background from this region. These constraints can reduce uncertainties on the background and can hence increase the sensitivity of the analysis.
The second advantage of a BDT-based analysis over a cut-based analysis is that a BDT is able to exploit the correlations between variables by using variables multiple times. This can be very useful if signal and background distributions overlap significantly in a set of variables, but have a different correlation for this set of variables allowing to separate signal and background.

[4] Assuming a branching ration of 1 % for both decays.

7.3.2 BDT Score as Final Discriminant

This analysis uses the BDT score as a final discriminant by performing a fit in this distribution to extract information on the signal. This strategy was selected in this analysis over the procedure of cutting on the BDT score and performing a fit with the remaining events in the distribution of a different discriminant (e.g. reconstructed Higgs boson mass) for two reasons.
The first and most important reason is that all events with a sufficient high BDT score are very signal-like meaning that the background has almost the same distribution as the signal in all variables that were either used directly in the training of the BDT or are correlated to these variables. This limits the fit's capability to discriminate between signal and background. The second reason for performing the fit in the distribution of the BDT score is that this allows to use all events in the fit. These additional events can constrain uncertainties of the background and improve therefore the fit's capability to extract information on the signal.
It was tested to cut on the BDT score and doing a fit in the reconstructed Higgs boson mass, both with and without using mass-correlated variables in the training of the BDT. Both options were outperformed by a fit in the distribution of the BDT score and were therefore not pursued in this analysis.

7.3.3 Binning of Final Discriminant

The selection of the binning of the BDT score used as the final discriminant has a large impact on the sensitivity of the analysis. Therefore, it is crucial to select a good binning to extract as much information on the signal from the dataset as possible. A good binning has to fulfill two criteria. First of all, it has to be selected such that it maximizes the sensitivity of the analysis extracted by the final fit. However, it has to be ensured that the sensitivity corresponds to a real physical sensitivity and is not caused by statistical downward fluctuations of the background estimates that are used as input parameters in the fitting procedure.
Due to the large computational effort it is not possible to perform a complete fit using the full distributions and considering all systematic uncertainties to calculate the expected significance as an estimator for the sensitivity for each tested binning. Therefore, the so called Asimov significance S_a is used as an estimator for the sensitivity. It is given by

$$S_a = \sqrt{2\left((s+b)\ln\left(1+\frac{s}{b}\right)-s\right)}, \tag{7.16}$$

using the number of signal events s and background events b.
Furthermore, most information on the signal is contained in the events that receive the highest BDT score, as these are the events that were classified as most signal-like by the BDT. This yields that the highest bin in the BDT score provides most information on the signal and will most likely be the major contributor to the overall significance of the analysis. Highest bin denotes the bin that is closest to a BDT score of 1. It was found that the highest bin contains more signal that any other bin, while providing the best signal to background ratio. Therefore, it was decided to optimize the binning by optimizing the bin width of the highest bin in the BDT score.
The bin width of the highest bin in the BDT score is optimized such that it gives the highest possible S_a under the constraint that each considered background has non-zero prediction for this bin with a relative statistical uncertainty of less than 100 %[5]. This constraint ensures that the estimate of all backgrounds is reliable and does not cause an unphysical sensitivity due to a downward fluctuation

[5] A relative uncertainty of 100 % might seem huge, but it has to be kept in mind that the fit performed to extract information on the signal considers these statistical uncertainties.

of the background estimate.
The optimized bin width of the highest bin in the BDT score is additionally used as the bin width for the other bins in the high BDT score region. The high BDT score region is defined as the region in the BDT score that contains a reasonable fraction of the signal. The background enriched low BDT score region is binned using a finer binning to exploit both the high statistics and the different shape of backgrounds in that region. The remaining medium BDT region, that contains neither a significant amount of signal, nor a large number of background events, is binned using a wider bin width. This is done to increase the number of events in these bins and avoid therefore that possible statistical fluctuations of the event yield in these bins affect the fit.
The previously discussed selection of the binning is overruled, if it suggests a binning that includes a bin with a zero prediction for any of the considered signal and background processes. In that scenario, the bin with of the affected bins is increased manually to avoid bins that do not have a prediction for all considered processes. This is done to avoid a bias in the fitting procedure that could be introduced by those bins as the fits performed in this analysis cannot consider processes with a zero prediction in the affected bins. This is critical since a zero prediction for a process in a bin does most likely not mean that this process has no contribution to this bin due to same physical reason, but is rather an artifact caused by the limited accuracy of the estimates.

7.3.4 Division into Four Final States

All fully leptonic final states are considered in this analysis meaning that all considered events contain exactly two leptons of opposite charge that can be either two muons, two electrons, or one muon and one electron. The analysis is divided in four different final states depending on the flavor of the leptons and their label. Two same flavor final states, ee and $\mu\mu$, are formed from all events containing two electrons and two muons, respectively. Two additional final states, μe and $e\mu$, are formed from the events containing two different flavor leptons. The lepton labeled as l_H is a muon (electron) for all events in the μe ($e\mu$) final state.
The events were split into the four different final states, as the considered final states have a significantly different background composition and are enriched in one of the two possible signals.
The $\mu\mu$ and ee final states are dominated by the overwhelming $Z \to \mu\mu$ and $Z \to ee$ background, respectively, yielding a very poor signal to background ratio in these final states. The different flavor final states have a much better signal to background ratio since they are barely affected by $Z \to \mu\mu$ and $Z \to ee$. The main background for the different flavor final states is the top quark background together with a significant contribution of $Z \to \tau\tau$ background. The $Z \to \tau\tau$ background has a much smaller contribution to the different flavor final states compared to the $Z \to \mu\mu$ and $Z \to ee$ contribution in the same flavor final states, since only about 6% of the τ-pairs originating from the Z decay into a final state contain exactly one electron and one muon.
The reason for splitting each of the same and different flavor final states in two final states is the possible existence of the two different signals, $H \to \tau\mu$ and $H \to \tau e$. The $H \to \tau\mu$ ($H \to \tau e$) signal will appear, up to detector effects, only in the $\mu\mu$ (ee) same flavor final state and up to mislabeling in the μe ($e\mu$) different flavor final state. Thus, the $\mu\mu$ and μe final state contain most of the $H \to \tau\mu$ signal events while the ee and $e\mu$ final states contain most of the $H \to \tau e$ allowing to combine those final states for a search for the corresponding signal. This is done in this analysis by performing a simultaneous fit for the $H \to \tau\mu$ ($H \to \tau e$) signal in the $\mu\mu$ and μe (ee and $e\mu$) final states described in section 9.3.
A contamination of the same flavor final state with the wrong[6] signal is very unlikely, since the

[6] Wrong Signal denotes here the $H \to \tau\mu$ ($H \to \tau e$) signal in the ee and $e\mu$ ($\mu\mu$ and μe) final state.

signal contains already one muon (electron) prior to the τ decay. This is different in the different flavor final states, where a mislabeling of the leptons yields automatically that a signal event will appear in the final state enriched with the other signal. This makes the wrong signal contamination in the different flavor final states significantly larger than in the same flavor final states, but the contamination with the wrong signal is still small, since the leptons are labeled correctly with an efficiency of roughly 90 %, as seen in section 6.4. The treatment of the wrong signal in the different flavor final states is discussed in section 7.4.5.

7.4 Selection of Different Flavor Final States

The μe and $e\mu$ final state are treated analogously. Events are split into two different regions, the signal region and the top quark control region. The latter includes all events containing at least one identified b-jet, while the signal region is formed from the remainder of events, as there are no b-quarks present at tree level for signal events in contrast to the b-quark originating from the t-quark decay. BDTs are trained separately in the signal region of the μe and $e\mu$ final state using only the VBF $H \to \tau\mu$ ($H \to \tau e$) sample as a signal in the BDTs of the μe ($e\mu$) final state. A detailed description of the training strategy for the BDTs in given in section 7.1.3. The distributions of the BDT score in signal and the event yield of the top quark control region are used for a fit extracting information on the signal described in chapter 9.

7.4.1 Event Selection for Different Flavor Final States

Table 7.1 lists the additional selections defining the different flavor final states of this analysis. These cuts are applied on all events passing the basic VBF selection described in section 5.3. The intention behind the selection of these cuts is explained in the following.
Cuts DF1 and DF2 define first of all the different flavor final states by demanding two leptons of different flavor and splitting additionally all events in the two different flavor final states, μe and $e\mu$, by demanding that l_H is a muon or electron, respectively. Cut DF3 is used to split the events of

	μe final state		**$e\mu$ final state**	
Cut name	μe Signal Region	μe top quark CR	$e\mu$ Signal Region	$e\mu$ top quark CR
DF1	l_H is muon	l_H is muon	l_H is electron	l_H is electron
DF2	l_τ is electron	l_τ is electron	l_τ is muon	l_τ is muon
DF3	$bTag = 0$	$bTag = 1$	$bTag = 0$	$bTag = 1$
DF4	$P_{l_H}^T > 20\,\text{GeV}$	$P_{l_H}^T > 20\,\text{GeV}$	$P_{l_H}^T > 20\,\text{GeV}$	$P_{l_H}^T > 20\,\text{GeV}$

Table 7.1: Additional cuts for different flavor final states defining the μe and $e\mu$ final state and their signal and top quark control region (CR). Events are kept, if the expression given in the table evaluates as true.

each final state into two different regions, the signal region and the top quark control region. The signal region is formed by all events containing no jet that was identified as a bTag. It contains most of the signal and has a reduced contribution of the top quark background. The top quark control region is in contrast to this enriched in the top quark background and contains only a small fraction of the signal. It is used to constrain the normalization of the top quark background in the signal region. The last cut, DF4, is of rather technical than physical nature. It was introduced to limit

effects of a potential mismodeling of the fake-lepton background by decreasing the contribution of the poorly modeled fake background that prefers small values of $P_{l_H}^T$ in contrast to the signal. Table 7.2 gives the normalization and scale factors used for the analysis of the different flavor final states. A description of the derivation of the embedding normalization and the fake-lepton scale factors can be found in section 4.2.2 and 4.2.3, respectively. The top quark background

Norm factor	μe **final state**	$e\mu$ **final state**
Top quark background normalization	1.076	1.105
Embedding normalization	0.157	0.158
Fake-lepton scale factor	0.212 ± 0.002	0.208 ± 0.002

Table 7.2: Scale and normalization factors used in the different flavor final states.

normalization is derived for each of the final states in their corresponding top quark control region. It is derived by scaling the top quark MC such that the sum of the scaled top quark MC and the other background estimates matches the observed data in the top quark control region. The top quark background normalization is then applied to the top quark control region and the signal region. This normalization of the top quark background is introduced at this stage to set the overall event yield of the top quark background to its best estimate. This allows to train the BDTs with the best estimate of the top quark background distribution. Later in the statistical evaluation, the normalization of the top quark background is treated as a free parameter in the performed fits.

	μe **final state**		$e\mu$ **final state**	
	μe Signal Region	μe Top Quark CR	$e\mu$ Signal Region	$e\mu$ Top Quark CR
top quark	684.2 ± 14.7	4151.2 ± 35.0	738.6 ± 15.6	4736.2 ± 37.9
$Z \to \tau\tau$ (Emb.)	265.0 ± 5.0	67.7 ± 2.7	402.7 ± 6.2	97.1 ± 3.2
WW/WZ/ZZ	159.8 ± 7.0	20.6 ± 2.7	190.6 ± 7.8	23.8 ± 3.0
Fake	95.0 ± 5.0	40.0 ± 4.7	313.9 ± 8.5	109.8 ± 6.5
$Z \to \mu\mu$	25.8 ± 5.8	2.7 ± 1.0	15.6 ± 1.6	1.5 ± 0.3
$Z \to ee$	1.2 ± 1.0	0.0 ± 0.0	2.4 ± 0.6	1.5 ± 0.9
VBF $H \to \tau\tau$	3.1 ± 0.0	0.4 ± 0.0	4.0 ± 0.0	0.5 ± 0.0
ggH $H \to \tau\tau$	2.8 ± 0.1	0.4 ± 0.0	3.7 ± 0.1	0.5 ± 0.0
VBF $H \to \tau\mu$	5.5 ± 0.1	0.7 ± 0.0	0.5 ± 0.0	0.1 ± 0.0
ggH $H \to \tau\mu$	5.1 ± 0.3	0.6 ± 0.1	0.5 ± 0.1	0.1 ± 0.1
VBF $H \to \tau e$	0.5 ± 0.0	0.1 ± 0.0	8.2 ± 0.1	1.0 ± 0.0
ggH $H \to \tau e$	0.8 ± 0.1	0.1 ± 0.1	7.9 ± 0.4	1.3 ± 0.2
Sum $H \to \tau\mu$	$\mathbf{10.7 \pm 0.3}$	$\mathbf{1.3 \pm 0.1}$	$\mathbf{1.0 \pm 0.1}$	$\mathbf{0.2 \pm 0.1}$
Sum $H \to \tau e$	$\mathbf{1.4 \pm 0.1}$	$\mathbf{0.2 \pm 0.1}$	$\mathbf{16.1 \pm 0.4}$	$\mathbf{2.3 \pm 0.2}$
Sum background	$\mathbf{1236.8 \pm 18.7}$	$\mathbf{4283.0 \pm 35.6}$	$\mathbf{1671.4 \pm 20.4}$	$\mathbf{4971.0 \pm 38.7}$
Observed data	$\mathbf{1212.0 \pm 34.8}$	$\mathbf{4283.0 \pm 65.4}$	$\mathbf{1684.0 \pm 41.0}$	$\mathbf{4971.0 \pm 70.5}$

Table 7.3: Expected event yields and observed data for the μe and ee final states in the signal and top quark control region (CR). Event yields for the LFV processes are calculated assuming a branching ratio of 1 % for $H \to \tau\mu$ and $H \to \tau e$, while all other event yield are calculated using the SM prediction. The $Z \to \tau\tau$ background estimate is done with die embedding method (Emb.). Only statistical uncertainties are listed.

Table 7.3 gives the expected event yields in the different regions of the different flavor final states of the processes considered in this analysis and compares it to the observed data. It can bee seen that

the top quark control region is indeed dominated by the top quark background with a purity of more than 95 % and contains a negligible amount of signal justifying the procedure of its normalization. The exact agreement of data and the sum of background in the two top quark control regions is artificial and caused by the normalization procedure of the top quark background.
Both signal regions are dominated by the top quark background and show a good agreement between the expected event yields and the observed data. The expected signal yield, that is calculated assuming a branching ratio of 1 %, is relatively small. The contamination with the wrong signal ($H \to \tau e$ in the μe final state and vice versa) is in both signal regions small compared to the proper signal. The strategy to deal with the wrong signal is discussed in section 7.4.5.
The overall event yield of the signal region of the $e\mu$ final state is higher compared to the one of the μe final state. This can be explained by the different trigger P^T thresholds of the leptons described in section 5.2. Muons have in the different flavor final states a smaller trigger threshold than electrons meaning that the trigger accepts events that contain muons with a P^T below the electron P^T threshold. The lepton labeling will assign in these events the muons most likely the label l_τ and the electron the label l_H, as the lepton labeling depends on the P^T difference of the leptons. This means that the $e\mu$ final state contains most of these events at the low P^T range of the muon. These events would not have be triggered if the flavor of the two leptons would be exchanged.
This effect is especially visible for backgrounds that have a high probability of producing one low P^T muon like the fake-lepton or $Z \to \tau\tau$. It can be seen in table 7.3 that these processes have a significantly larger event yield in the $e\mu$ final state, while processes that produce mainly high P^T objects like the top quark background are only affected weakly by this effect.
This effects the signal yield as well, as l_τ is most likely low in P^T, causing that more $H \to \tau e$ events, where the muon is most likely labeled as l_τ, will be accepted by the trigger than $H \to \tau\mu$ events, where the electron is more likely labeled as l_τ. This causes a higher event yield for the $H \to \tau e$ signal, as seen in table 7.3.

7.4.2 BDT Training

All events from the μe and $e\mu$ signal regions were used in the training of BDTs in the corresponding final states. Table 7.4 gives the variables that were used in the training of the BDTs for each

Variable	**Rank in μe final state**	**Rank in $e\mu$ final state**
m_{coll}	1	1
$m^T(E^T_{\text{miss}}, l_\tau)$	2	2
m_{ll}	3	3
m_{jj}	4	4
x_{l_H}	not considered	5
$\Delta\eta_{jj}$	5	6
m^T_{ll}	6	not considered
$P^T_{l_H}$	7	7
E^T_{miss}	8	not considerd
$S(d0_{l_\tau})$	9	not considered

Table 7.4: Variables used in the training of the BDTs in the μe and $e\mu$ final state with their ranking.

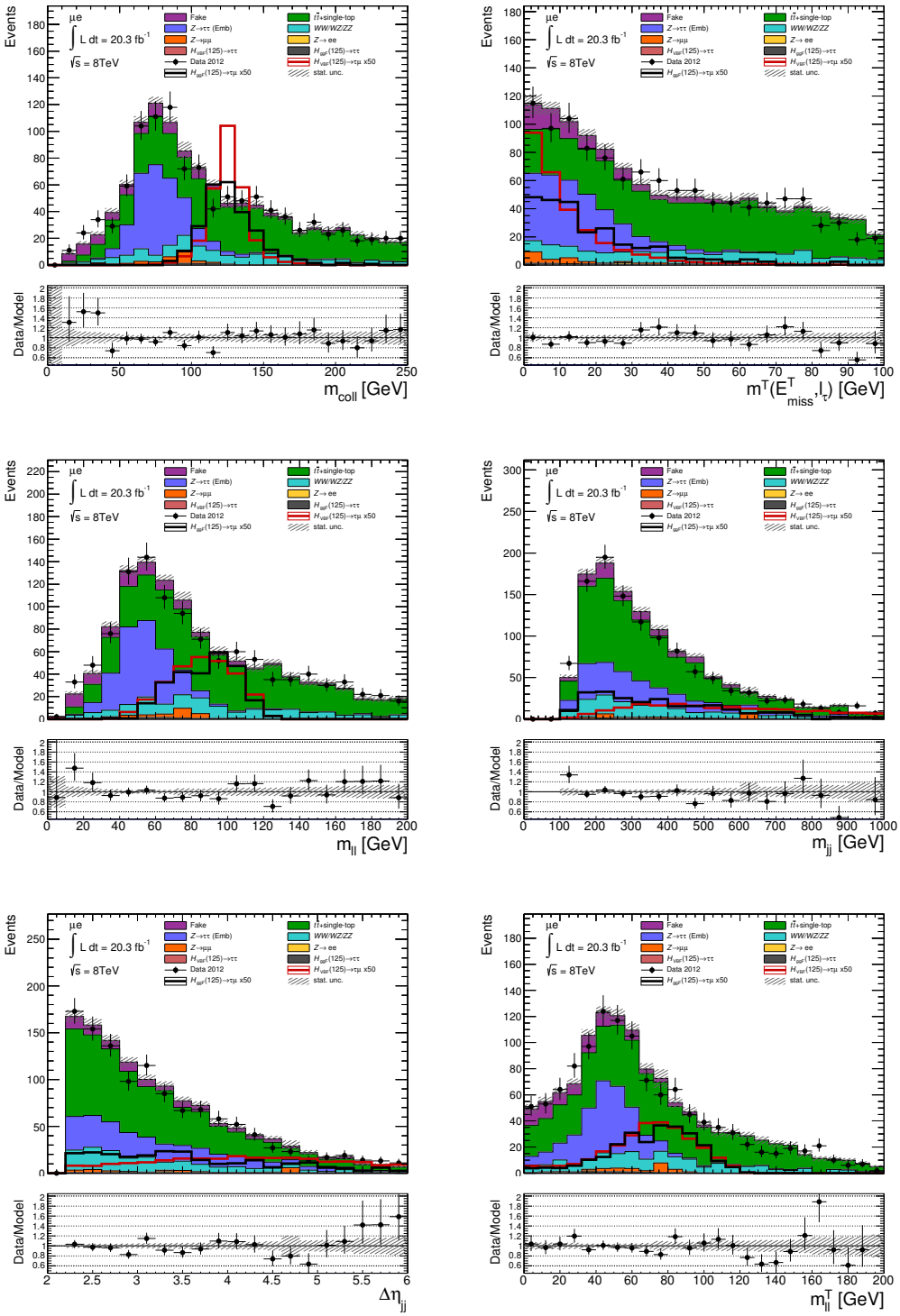


Figure 7.6: Distributions of input variables of the BDTs in the μe final state. The definitions of the variables can be found in section 7.2. All plots show in the upper box the stacked background contributions and the observed data (black dots). Additionally, the $H \to \tau\mu$ signal that is enhanced by a factor of 50 is shown as a red (black) line for the VBF (ggH) production process. The lower box gives the ratio of the observed data and the sum of all background models. All shown error bars and error bands correspond only to statistical uncertainties.

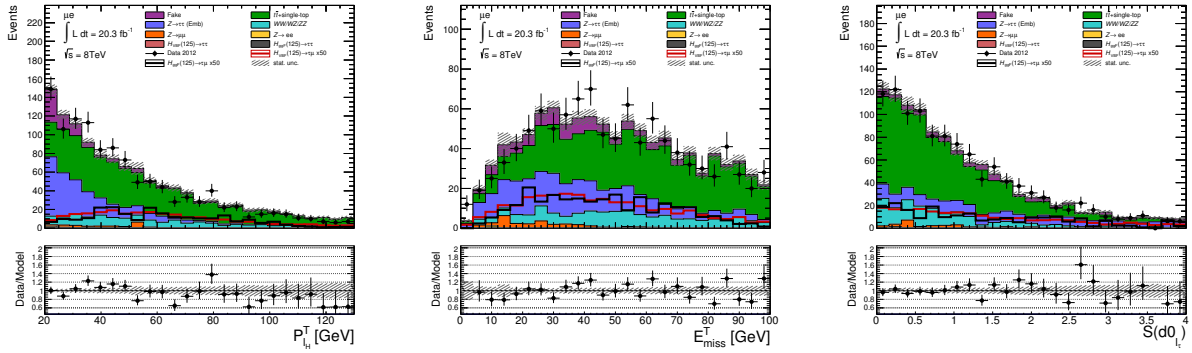


Figure 7.7: Distributions of input variables of the BDTs in the μe final state. The definitions of the variables can be found in section 7.2. All plots show in the upper box the stacked background contributions and the observed data (black dots). Additionally, the $H \to \tau\mu$ signal that is enhanced by a factor of 50 is shown as a red (black) line for the VBF (ggH) production process. The lower box gives the ratio of the observed data and the sum of all background models. All shown error bars and error bands correspond only to statistical uncertainties.

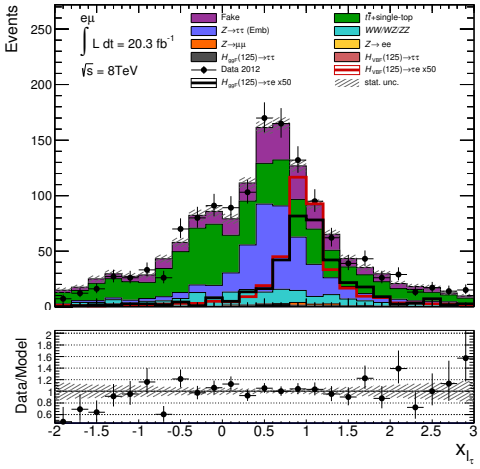


Figure 7.8: Distribution of the input variable x_{l_τ} for the BDTs in the $e\mu$ final state. The definition of the variable can be found in section 7.2. This plot shows in the upper box the stacked background contributions and the observed data (black dots). Additionally, the $H \to \tau e$ signal that is enhanced by a factor of 50 is shown as a red (black) line for the VBF (ggH) production process. The lower box gives the ratio of the observed data and the sum of all background models. All shown error bars and error bands correspond only to statistical uncertainties.

different flavor final states and their ranking by the BDTs. A high rank (small number) corresponds to a high usage of this variable by the BDTs. It can be seen that both BDTs make use of a similar set of variables with a very similar ranking pronouncing how similar both same flavor final states are. Other variables were tested but found to contribute insignificantly and were therefore neglected. This procedure favors a slightly different set of variables for the two final states, since the final states have a different background composition.

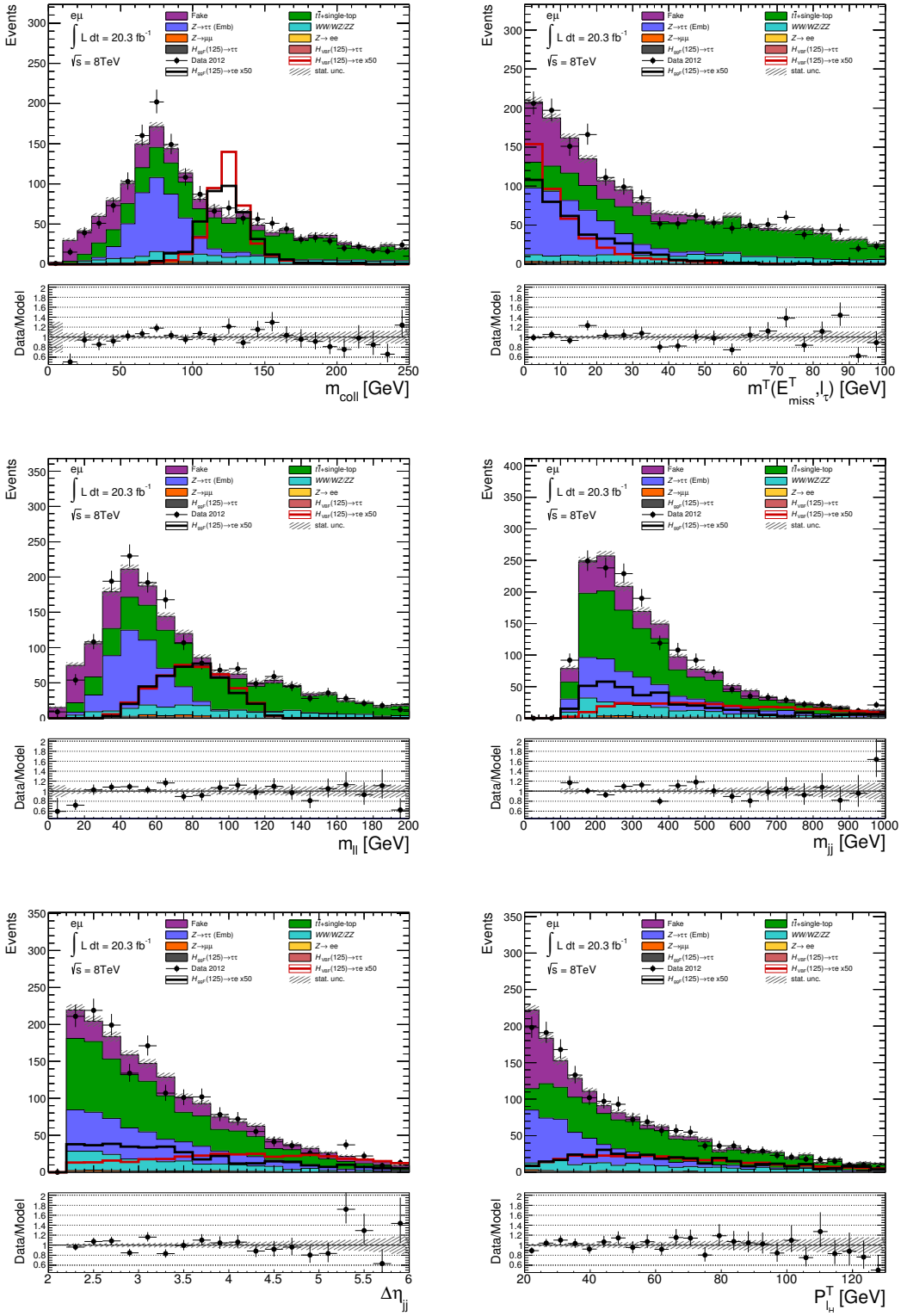
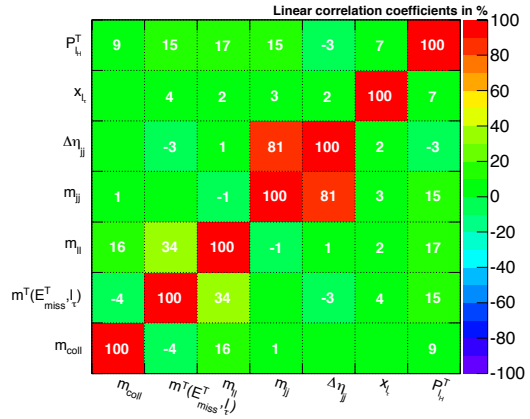
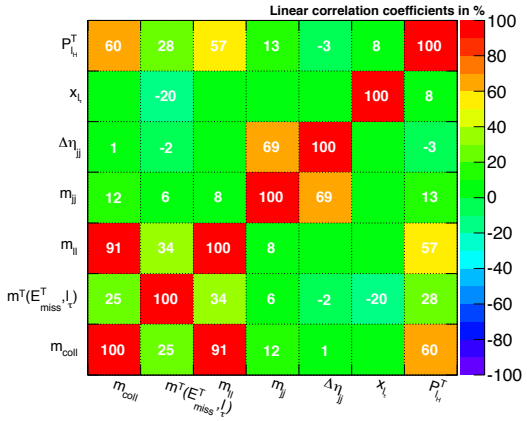


Figure 7.9: Distributions of input variables of the BDTs in the $e\mu$ final state. All plots show in the upper box the stacked background contributions and the observed data (black dots). Additionally, the $H \rightarrow \tau e$ signal that is enhanced by a factor of 50 is shown as a red (black) line for the VBF (ggH) production process. The lower box gives the ratio of the observed data and the sum of all background models. All shown error bars and error bands correspond only to statistical uncertainties.

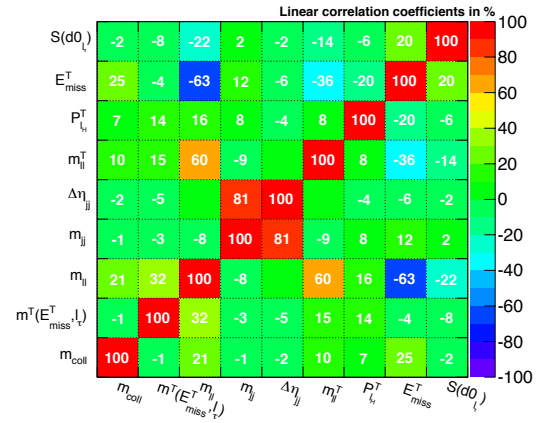
The distributions of the input variables of the μe final state are shown in figure 7.6 and figure 7.7. The corresponding distributions of the input variables of the $e\mu$ final state are shown in figure 7.9 and figure 7.8. All variables show a sufficient agreement between data and MC prediction justifying their use as input variables for the BDTs. It can be seen that the variables describing masses like m_{coll} and $m^T(E^T_{\text{miss}}, l_\tau)$ show good separation between signal and background and it is therefore not surprising that they are ranked as the most important variables by the BDTs.



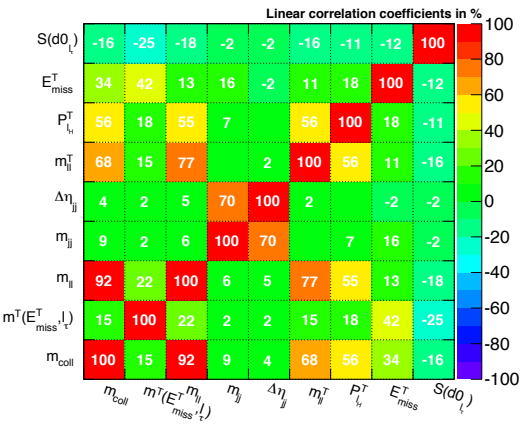
(a) VBF $H \to \tau e$ in the $e\mu$ final state



(b) weighted sum of all background processes in the $e\mu$ final state



(c) VBF $H \to \tau\mu$ in the μe final state



(d) weighted sum of all background processes in the μe final state

Figure 7.10: Plots of the linear correlation coefficient of the input variables of the BDTs for the $e\mu$ (μe) final state in the top (bottom). The plots on the left hand side are for the signal used in the BDT training, while the plots on the right correspond to the weighted sum of all background samples used in the BDT training.

The correlation of the input variables are visualized in figure 7.10 showing the linear correlation coefficient. These plots show that the correlations of the input variables are different for signal and background, as for example for the variables m_{ll} and m_{coll}. They are strongly correlated in the background, while they only show a lose correlation for the signal. A BDT is able to exploit such correlations in contrast to a cut based analysis.

7.4.3 Binning

The optimization of the binning is carried out according to the strategy described in section 7.3.3. However, the $Z \to ee$ and $Z \to \mu\mu$ background samples had to be added and treated as a single background, since only few of the $Z \to ee$ MC events pass the selection leaving a wide range of the BDT score without a prediction of this specific background. Not adding these backgrounds would therefore yield very large bins to ensure that each background has a predicted event yield in each bin decreasing the sensitivity significantly. This procedure is justified by the relatively small contributions of both backgrounds in this final state and the fact that both backgrounds behave very similarly. The $Z \to ee$ and $Z \to \mu\mu$ background are as well summed up and treated as a single background in the statistical evaluation of the different flavor final states.

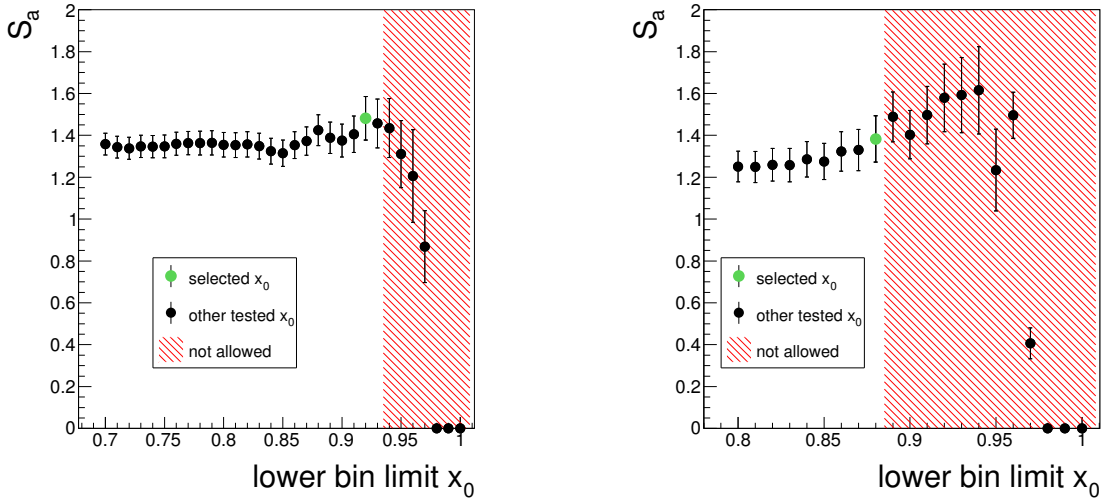


Figure 7.11: Asimov significance S_a of the highest bin of the BDT score that contains all events in the interval $[x_0, 1]$ depending on the lower bin bound x_0 on the left (right) for $e\mu$ (μe) final state. The Asimov significance corresponding to the selected x_0 is marked in green, while the Asimov significance of all other tested x_0 are shown in black. The selection of a x_0 lying in the red shaded area is not allowed by the requirement on the background predictions. The shown uncertainties are inferred from the statistical uncertainties of the signal and background predictions.

The binning in the signal regions is optimized with respect to the highest bin in the BDT score containing all events with a BDT score in the interval $[x_0, 1]$ using the lower bound of the bin x_0. Multiple values of x_0 were tested with an increment of 0.01 in the BDT score using the optimized binning. Figure 7.11 shows the tested x_0 and the corresponding Asimov significance of the highest bin of the BDT score. The selected values of x_0 were 0.92 and 0.88 for the $e\mu$ and the μe final state, respectively.
The top quark control regions were selected to consist of only one bin, since they are only used to constrain the overall normalization of the top quark background. Distributions of the BDT score with the optimized binning in the signal region are shown in the next section.

7.4.4 BDT Score

The BDT score for the signal and top quark control region of both different flavor final states is shown figure 7.12 in the binning that was selected in the previous section. First of all, it can be seen that the BDTs are able to achieve a good separation between signal and background enriching the low BDT score regions of the signal regions with background, while the high BDT score region is enriched with signal.
None of the distributions shows a significant deviation between the sum of the background estimates and the observed data. The perfect agreement in the top quark control region is however an artifact of the normalization procedure of the top quark background. The good agreement between the sum of the background estimates and the observed data in both signal regions allows to draw two conclusions. First, no obvious signal is present as this would yield an excess of data in the high BDT score region. The highest bin of the BDT score of the μe final state shows nevertheless a mild excess of observed data events compared to the background prediction. This mild excess is however not statistically significant. The second conclusion is that the BDT scores of both final states are modeled well by the background estimates. This is important as the BDT score will be used as the final discriminant in the fits performed in the statistical evaluation of this analysis.
Figure 7.13 shows the comparison of the shape of the BDT score between all considered backgrounds and the added signal (ggH +VBF). First of all, it can be seen that the signal shows a peak structure for high BDT scores close to 1, while all backgrounds peak at a BDT score close to -1. This is an important feature of the distribution of the final discriminant for the fitting procedure, as this means that none of the backgrounds will be able to fake a peak at a BDT score close to 1 that could be misinterpreted as a signal. This holds also true for the $H \to \tau\tau$ background meaning that the BDTs are able to discriminate well between the LFV Higgs boson decays, $H \to \tau\mu$ and $H \to \tau e$, and the SM Higgs boson decay $H \to \tau\tau$ preventing that it can be misidentified as a signal.
These plots highlight additionally the justification for the fine binning at a BDT score close to -1. The two main backgrounds, top quark and $Z \to \tau\tau$, show a different peak structure that is only resolved due to the fine binning. The $Z \to \tau\tau$ background peaks at the lowest bin of the BDT score, while the top quark background peaks in the second to last bin and has a relatively small contribution in the lowest bin. This can help the fit to constrain the two backgrounds in contrast to the situation where both backgrounds peak in the same bin. Within this situation, the fit would be free to scale one background up and the other down while keeping the expected event yields in the bin constant. This freedom of the fit is obviously limited if both backgrounds show a smaller overlap.

7.4.5 Handling of Wrong Signal

This analysis considers two different flavor final states μe and $e\mu$. They are enriched in the $H \to \tau\mu$ and $H \to \tau e$ signal, respectively, but not pure with respect to the other signal, denoted as the wrong signal in the following.
A wrong labeling of the leptons means that a signal event does not appear in its dedicated final state, but in the final state dedicated to the other signal. Table 7.5 shows the expected event yield of both signals in both final states. It can be seen that the expected event yields of the wrong signal are in both final states about one order of magnitude smaller that the expected event yields of the signal. Table 7.5 gives additionally the ratio of events that show up in the right final state, that is defined as the ratio of the expected yield of the $H \to \tau\mu$ ($H \to \tau e$) signal in the μe ($e\mu$) final states and the total expected yield of this processes. This ratio agrees well with the matching ratio of the lepton labeling given in section 6.4.

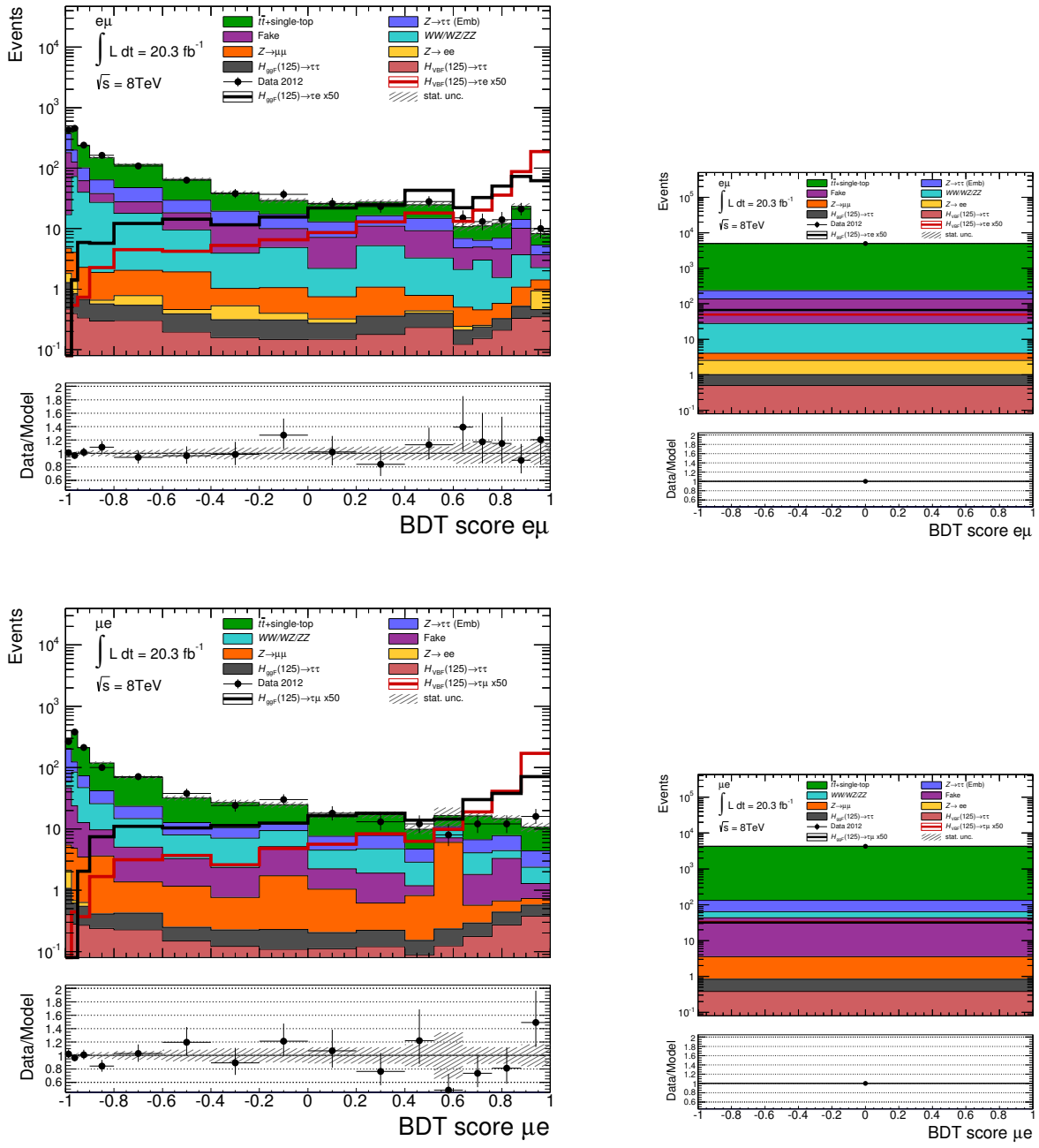


Figure 7.12: Distribution of the BDT score in the signal region (left) and composition of the top quark control region (right) of both different flavor final states. The top (bottom) row shows the $e\mu$ (μe) final state. All shown uncertainties correspond to the statistical uncertainties only.

The shape of the BDT score of both signals is compared in figure 7.14 for both different flavor final states. The wrong signal shows clearly a different shape compared to the signal. This is due to the fact that a right labeling is crucial in the calculations of variables like m_{coll} and $m^T(E^T_{\text{miss}}, l_\tau)$, since their calculation relies on the properties of l_H and l_τ. A wrong labeling yields therefore that

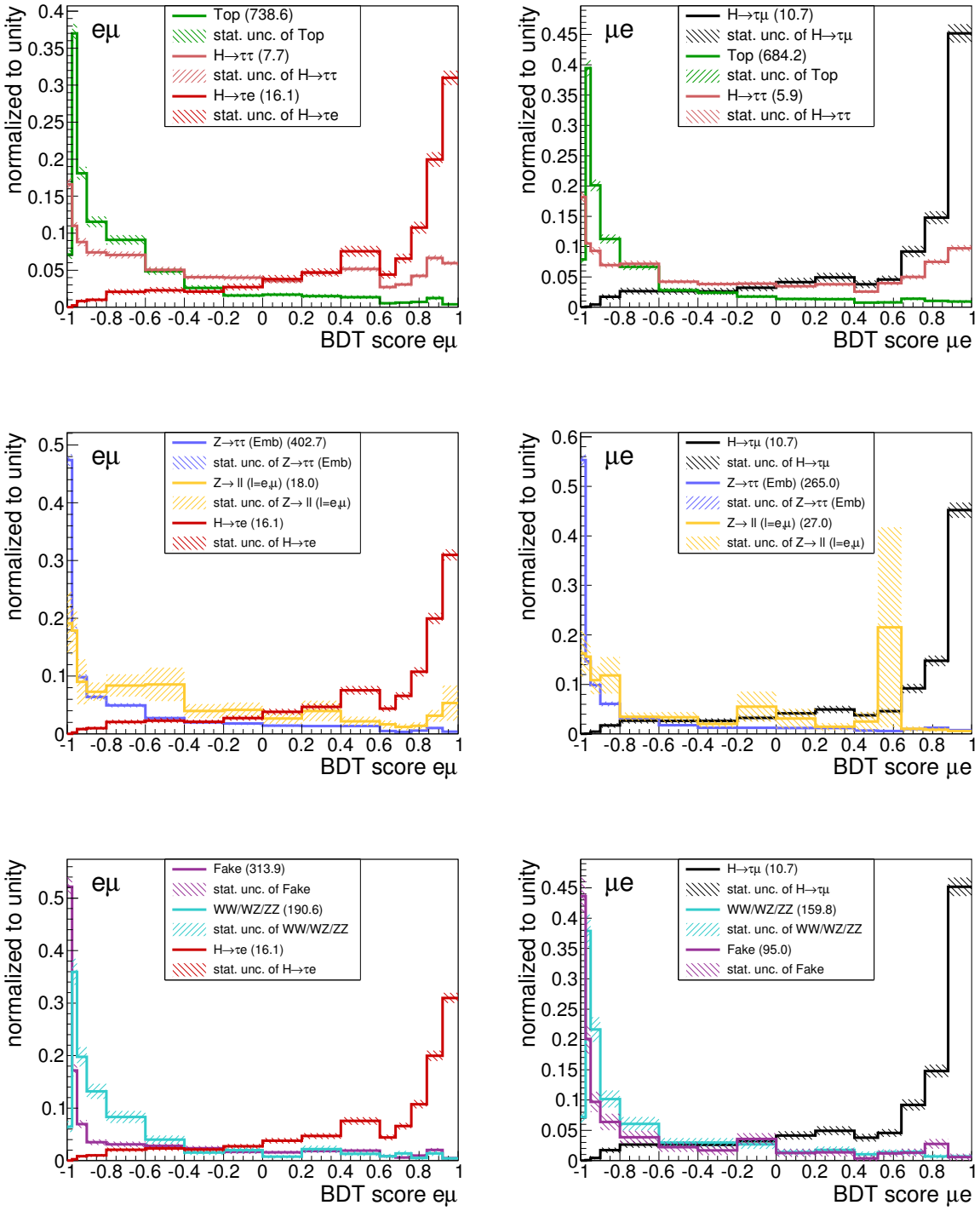


Figure 7.13: The left (right) plots shows the shape of the distribution of signal and background for the signal region of the $e\mu$ (μe) final states in the corresponding BDT score. The contributions of the ggH and VBF production modes of the various Higgs decays have been added up. The $Z \to ee$ and $Z \to \mu\mu$ backgrounds have been added up as well and are denoted as $Z \to ll$. All shown distributions are normalized to unity and the given uncertainties are statistical. The number stated in the brackets corresponds to the expected event yield.

Signal	μe and $e\mu$ SR	$e\mu$ SR	μe SR	Ratio in right final state
$H \to \tau\mu$	11.6 ± 0.3	0.99 ± 0.09	10.7 ± 0.3	0.91 ± 0.04
$H \to \tau e$	17.4 ± 0.4	16.1 ± 0.4	1.38 ± 0.12	0.92 ± 0.03

Table 7.5: Comparison of expected event yields of both considered signals in the signal region (SR) of the μe and $e\mu$ final state assuming a branching ratio of 1 %. The last column gives the fraction of signal events that appear in the signal region of the right final state and the sum of events of that signal appearing in the signal region of any (μe or $e\mu$) final state. All stated uncertainties correspond to statistical uncertainties.

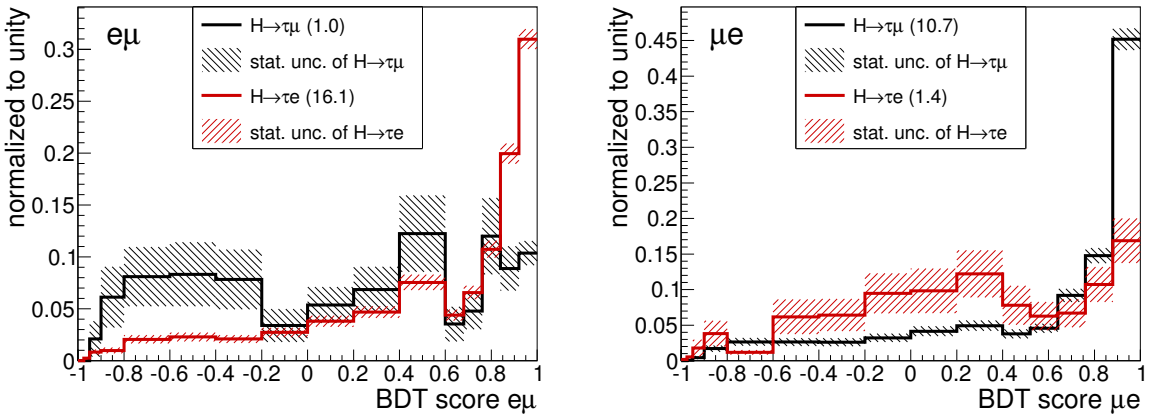


Figure 7.14: Theses plots show the shape of the wrong signal in the signal region of both different flavor final states, left (right) for the $e\mu$ (μe) final state. All shown distributions are normalized to unity and the given uncertainties are purely statistical.

a signal event behaves on average more background-like in these variables and receives therefore a rather low BDT score. The different shape of the wrong signal makes it unlikely that it will be able to fake a signal in the fit.
Therefore, this analysis neglects the wrong signal in the fitting procedure described in chapter 9. There is an additional physical reason justifying this neglect of the wrong signal. The indirect limits, presented in section 2.2.2, do not allow both signals to be large and therefore observable simultaneously, justifying to consider only one signal at a time.

7.5 Selection of Same Flavor Final States

The ee and $\mu\mu$ final states are treated analogously in this analysis. The main challenge in these same flavor final states is dealing with the large $Z \to ll$ (l=e,μ) background. Following the same strategy as in the different flavor final states was found to be unfavorable, as the BDT is overwhelmed by the dominant $Z \to ll$ background. This means that a BDT trained at this stage is not sensitive to any other backgrounds than $Z \to ll$, since the contribution of the other backgrounds to the total background is too small to make a difference. The result of this is that the BDT is only able to discriminate between signal and the $Z \to ll$ background yielding a poor suppression of the other backgrounds. This is problematic for two reasons. First, this yields a poor signal to background

ratio in the region of the BDT score containing the signal, because backgrounds other than $Z \to ll$ are not suppressed. Second, it can cause problems in the fitting procedure since the unsuppressed backgrounds tend to peak at the same position in the BDT score as the signal does. This can affect the fit's capability to extract the signal from data. Both issues result in a non optimal sensitivity of the analysis. This issue is dealt with training two sets of BDTs [7]. In both sets of BDTs, BDTs are trained separately in the signal region of the $\mu\mu$ and ee final state using only the VBF $H \to \tau\mu$ ($H \to \tau e$) sample as a signal in the BDTs of the $\mu\mu$ (ee) final state. A detailed description of the training strategy for the BDTs in given in section 7.1.3.
The first set of BDTs, denoted from here on as BDTs_1, are trained on all events of the same flavor final states in each of the $\mu\mu$ and ee final states separately. The BDT score of the BDTs_1 is then cut on to suppress the $Z \to ll$ background. Events that receive a higher BDT score than the selected cut value form the signal region, while all other events form a $Z \to ll$ control region. The $Z \to ll$ control region is used to infer a normalization for the $Z \to ll$ background.
The events in the signal region are then used to train a second set of BDTs, denoted as BDTs_2, that are sensitive to all backgrounds due to the significantly reduced contribution of the $Z \to ll$ background. The BDTs_2 are then evaluated to calculate BDT scores in the signal region of the corresponding final states. The distribution of this BDT score in the signal regions and the event yield of the $Z \to ll$ control regions is finally used for the fit extracting the signal, described in chapter 9. The analysis of the same flavor final states is discussed in the following in detail.

7.5.1 First Set of BDTs

Event Selection

Table 7.6 list the cuts used in the first part of the analysis of the same flavor final states. These cuts are applied on top of the basic VBF selection described in section 5.3.

Cut name	ee final state	$\mu\mu$ final state
SF1	l_H is electron	l_H is muon
SF2	l_τ is electron	l_τ is muon
SF3	$bTag = 0$	$bTag = 0$
SF4	$m_{ll} > 30\,\text{GeV}$	$m_{ll} > 30\,\text{GeV}$
SF5	$m_{\text{coll}} > 80\,\text{GeV}$	$m_{\text{coll}} > 80\,\text{GeV}$

Table 7.6: Additional cuts for the same flavor final states. Events are kept, if the expression given in the table evaluates as true.

Cuts SF1 and SF2 define the ee and $\mu\mu$ final state by demanding that both leptons, l_H and l_τ, are electrons or muons, respectively. The cut SF3 accepts only events that do not contain a jet identified as a b-jet. This cut was introduced to suppress the contribution of events containing b-jets, as the $Z \to ee$ and $Z \to \mu\mu$ MC samples used in this analysis do not contain any matrix elements for the heavy quarks[8]. Not suppressing events with an identified b-jet would therefore cause an observable difference between the MC prediction and the data. This cut also decreases

[7] The SM $H \to \tau\tau$ analysis of the ATLAS collaboration [37] dealt with this issue by introducing a very effective upper cut on m_{ll} to suppress the $Z \to ll$ background. This is unfortunately not possible in this analysis, as the LFV signal considered in this analysis has a higher di-lepton mass compared to $H \to \tau\tau$, causing a significant overlap of signal and the $Z \to ll$ in this variable.

[8] MC samples of $Z \to ll$ background including the matrix elements for the heavy quarks are in principle available, but it was beyond the timescale of this masters project to include them into the analysis framework.

the contribution of the top quark background.
Unfortunately, it is not possible to use the events that contain an identified b-jet to form a meaningful top quark control region as in the different flavor final states. This is due to the overwhelming event yield of the $Z \to ll$ background in the same flavor final stats in combination with the missing MC prediction for the production of the $Z \to ll$ background in association with heavy quarks. This means that such a top quark control region has a large contribution from the $Z \to ll$ background that is modeled very poorly in that specific phase space. This causes such a top control region to show a significant deficit in MC prediction compared to the observed data, that is not caused by a wrong normalization of the top quark background. Due to this and the fact that the top quark background is only a minor contributor to the total background, it was decided not to introduce a top quark control region for the same flavor final states.
The last two cuts, SF4 and SF5, are introduced to exclude a certain part of the phase space from the analysis, that is modeled poorly by the MC samples. This is not hurting the sensitivity, as the signal prefers larger values for both variables than the threshold values of the cuts.
Table 7.7 gives the applied normalization and scale factors. The normalization of the $Z \to ee$ and $Z \to \mu\mu$ background is inferred from all events in the corresponding final state, that are included in the Z-peak region, defined by $80\,\text{GeV} < m_{ll} < 100\,\text{GeV}$. The Z peak is very pure with respect

Norm factor	***ee* final state**	**$\mu\mu$ final state**
$Z \to ee$ normalization	1.120	n.a
$Z \to \mu\mu$ normalization	n.a.	1.000
Embedding normalization	0.192	0.136
Fake scale factor	0.22 ± 0.03	0.337 ± 0.017

Table 7.7: Scale and normalization factors used in the same flavor final states. The normalization factors for the $Z \to \mu\mu$ and $Z \to ee$ are only used for the training and evaluation of BDTs_1 and will be replaced by normalization factors from the proper Z control regions, once they are at hand.

to the other backgrounds and the signal, as can bee seen in table 7.8 justifying the use of it for the calculation of the norm factors. The norm factor is calculated by scaling the $Z \to ll$ background in the Z-peak region such, that the expected event yield matches the observed data.
Table 7.8 shows the event yields for both final states, once for all events contained in the final state and once for the Z-peak region. It can be seen that both final states are overwhelmed by $Z \to ll$ background and contain an expected signal yield of the same order of magnitude as the different flavor final states. The $\mu\mu$ final state has a higher event yield than the ee final state due to different trigger P^T thresholds and efficiencies. The overall expected event yields are in a good agreement with the observed data.

BDT Training

Individual BDTs are trained with all events contained in each individual final state, starting with the whole set of variables except $d0_{l_\tau}$ and $S(d0_{l_\tau})$. These variables were excluded from the training of the BDTs, as they were found to be modeled unsatisfactory in the MC samples at this stage. The final set of variables used for the training of BDTs_1 is listed in table 7.9 for both final states. Additionally, the rank that was assigned to each variable by the BDTs is given. A high rank (small number) corresponds to a high usage of that variable in the BDTs.
The fact that BDTs of both final states prefer to use the same set of variables is expected, since the $Z \to \mu\mu$ and $Z \to ee$ behave very similarly and the composition of the backgrounds are almost

	ee final state		**$\mu\mu$ final state**	
	all events	Z-peak region	all events	Z-peak- region
$Z \to \mu\mu$	0.0 ± 0.0	0.0 ± 0.0	89668.4 ± 205.7	80791.8 ± 195.3
$Z \to ee$	63105.5 ± 180.5	57485.5 ± 173.3	0.0 ± 0.0	0.0 ± 0.0
top quark	471.4 ± 11.5	84.5 ± 5.0	623.2 ± 13.8	119.6 ± 6.2
$Z \to \tau\tau$ (EMB)	108.2 ± 3.7	6.3 ± 1.0	220.4 ± 4.2	11.7 ± 1.1
WW/WZ/ZZ	489.4 ± 10.2	354.3 ± 8.3	660.2 ± 12.3	478.5 ± 9.8
Fake-leptons	189.4 ± 14.4	60.0 ± 12.6	185.6 ± 18.9	59.8 ± 17.6
VBF $H \to \tau\tau$	2.5 ± 0.0	0.4 ± 0.0	4.1 ± 0.0	0.5 ± 0.0
ggH $H \to \tau\tau$	1.9 ± 0.1	0.3 ± 0.0	3.5 ± 0.1	0.5 ± 0.0
VBF $H \to \tau\mu$	0.0 ± 0.0	0.0 ± 0.0	9.4 ± 0.1	3.3 ± 0.1
ggH $H \to \tau\mu$	0.0 ± 0.0	0.0 ± 0.0	9.4 ± 0.4	3.9 ± 0.2
VBF $H \to \tau e$	5.9 ± 0.1	2.3 ± 0.1	0.0 ± 0.0	0.0 ± 0.0
ggH $H \to \tau e$	5.8 ± 0.3	2.0 ± 0.2	0.0 ± 0.0	0.0 ± 0.0
Sum $H \to \tau\mu$	$\mathbf{0.0 \pm 0.0}$	$\mathbf{0.0 \pm 0.0}$	$\mathbf{18.9 \pm 0.4}$	$\mathbf{7.1 \pm 0.3}$
Sum $H \to \tau e$	$\mathbf{11.7 \pm 0.3}$	$\mathbf{4.3 \pm 0.2}$	$\mathbf{0.0 \pm 0.0}$	$\mathbf{0.0 \pm 0.0}$
Sum background	$\mathbf{64368.3 \pm 181.8}$	$\mathbf{57991.3 \pm 174.0}$	$\mathbf{91365.4 \pm 207.4}$	$\mathbf{81462.4 \pm 196.5}$
Observed data	$\mathbf{64853.0 \pm 254.7}$	$\mathbf{57991.0 \pm 240.8}$	$\mathbf{91182.0 \pm 302.0}$	$\mathbf{81462.0 \pm 285.4}$

Table 7.8: Event yields in the ee and $\mu\mu$ final state in the signal region and at the Z peak of the signal region, that contains all events of the signal region with $80\,\text{GeV} < m_{ll} < 100\,\text{GeV}$.

Variable	**Rank in ee final state**	**Rank in $\mu\mu$ final state**
m_{coll}	1	1
m_{jj}	2	4
m_{ll}	3	2
x	4	3
$\Delta\phi(l_\tau, E^T_{\text{miss}})$	5	5
$\Delta\eta_{jj}$	6	6

Table 7.9: Variables and their ranking used in the training of the first BDTs in the ee and $\mu\mu$ final state.

identical in both final states.
Figure 7.15 and 7.16 show the distribution of all used input variables in the ee and $\mu\mu$ final state, respectively. All variables show a good separation between signal and background and a sufficiently good modeling to justify their usage in the BDTs.
The correlations between the input variables are shown in figure 7.17. It can be seen that signal and background show different correlations in multiple variable pairs that can be exploited by the BDTs. It is expected that the correlation matrices of both final states are very similar, but the the fact that they are identical is a coincidence.

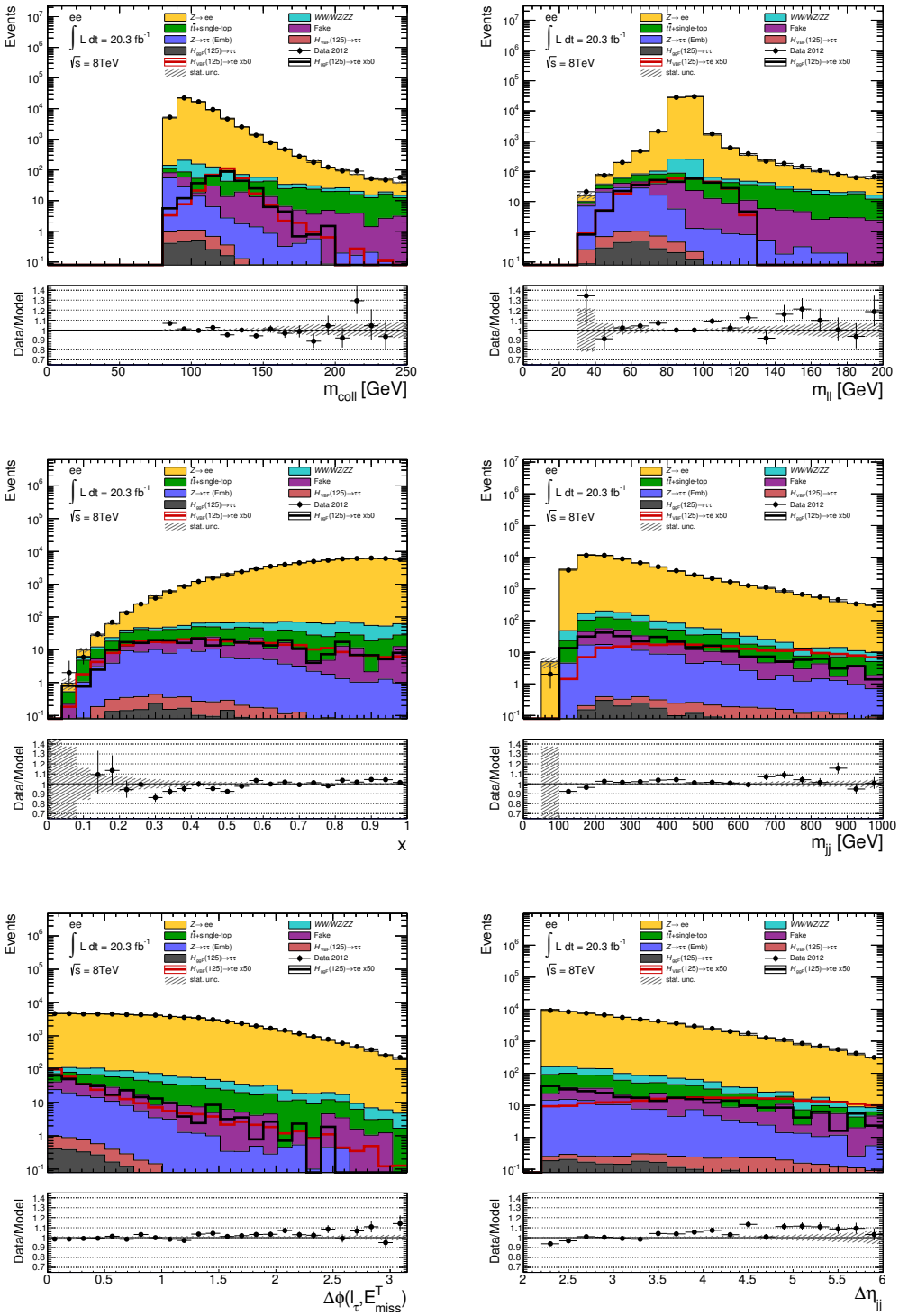


Figure 7.15: Input variables for the training of BDTs_1 in the *ee* final state. The definitions of the variables can be found in section 7.2. All plots show in the upper box the stacked background contributions and the observed data (black dots). Additionally, the $H \to \tau e$ signal that is enhanced by a factor of 50 is shown as a red (black) line for the VBF (ggH) production process. The lower box gives the ratio of the observed data and the sum of all background models. All shown uncertainties correspond only to the statistical uncertainties.

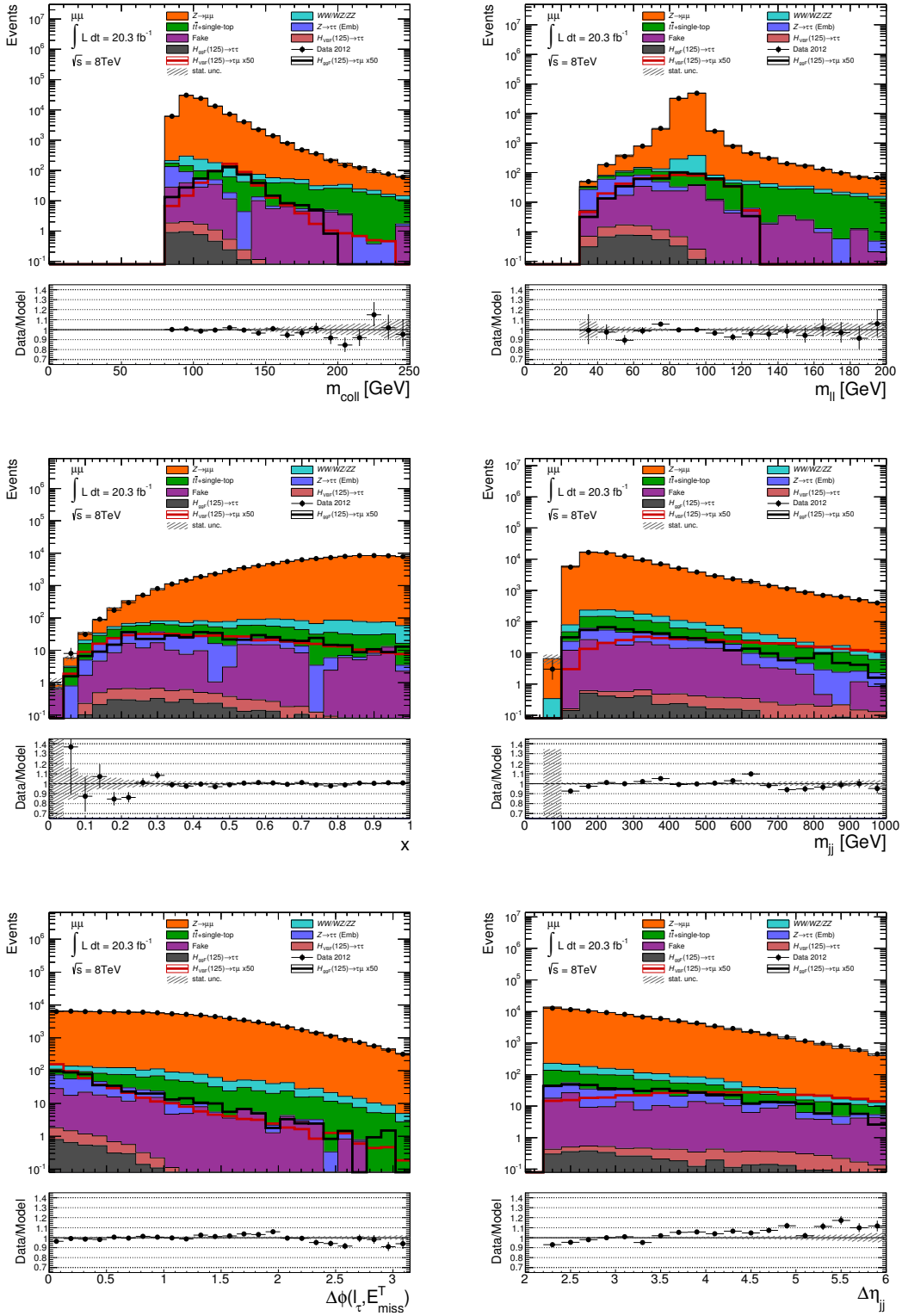


Figure 7.16: Input variables for the training of BDTs_1 in the $\mu\mu$ final state. The definitions of the variables can be found in section 7.2. All plots show in the upper box the stacked background contributions and the observed data (black dots). Additionally, the $H \to \tau\mu$ signal that is enhanced by a factor of 50 is shown as a red (black) line for the VBF (ggH) production process. The lower box gives the ratio of the observed data and the sum of all background models. All shown uncertainties correspond only to statistical uncertainties.

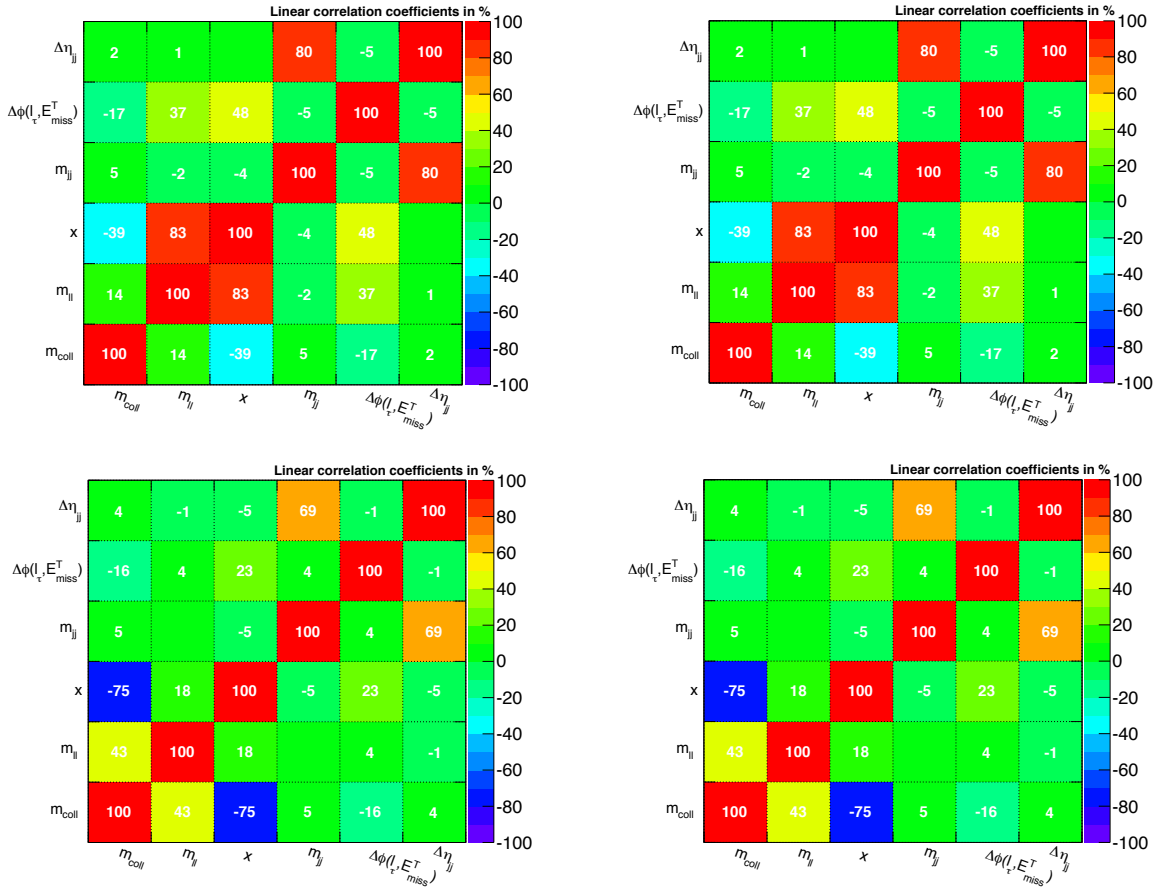


Figure 7.17: The linear correlation coefficient of the input variables of the first set of BDTs for the ee ($\mu\mu$) final state on the left (right) hand side. The plots on top are for the signal used in the BDT training, while the plots on bottom correspond to the weighted sum of all background samples used in the BDT training.

BDT Score of First Set of BDTs

Figure 7.18 shows the BDT score of both final states. First of all it can be seen that the BDT is not able to make use of the whole range of the BDT score given by $[-1, 1]$. This is due to the fact that the BDT is not able to discriminate well between signal and background. Each decision tree classifies an event either as signal (=1) or background (=-1). The BDT score is the weighted average of theses responses of all individual decision trees. This means that an event receives only a BDT score close to 1 (-1) if almost all decision trees classify the event as signal (background). This means that the BDTs shown in this section are not able to classify events with very high accuracy, as they do not assign BDT scores close to 1 or -1.
The BDTs manage nevertheless to gain a certain separation between signal and background, as the signal peaks for both final states in the bin with the highest assigned BDT score, while the sum of all backgrounds peaks at the lower range of the BDT score. The BDT score is furthermore described reasonably by the background estimates, as the distribution of the BDT score of both final states show no significant deviation between the observed data and the background estimates. The binning of the BDT scores given in this section is not optimized, since this analysis does not

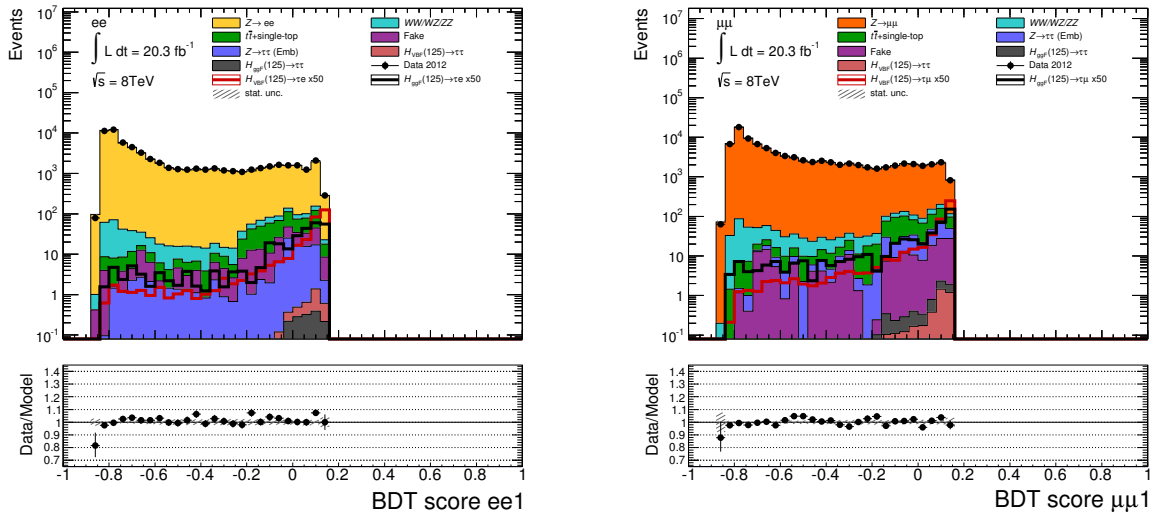


Figure 7.18: BDT score of the first set of BDTs in the ee final state (left) and in the $\mu\mu$ final state (right). The shown signal is enhanced by a factor of 50 and the given uncertainties correspond only to the statistical uncertainties.

use the shown BDT scores as an input for a statistical evaluation.

Motivation for Using Second BDTs

Figure 7.18 shows additionally two worrisome features of the BDT sore of the first BDTs in the same flavor final states. It is quite obvious that the signal to background ratio is very low, even when considering only the highest bins in the BDT score, as the signal that is enhanced by a factor of 50 does not overshoot the background estimates in the High BDT score region. It can be seen as well, that some of the backgrounds peak at high values of the BDT score. This problem is especially visible in figure 7.19 that shows the normalized distributions of the BDT score for signal and background processes. The BDTs of both final states are not able to discriminate between the LFV Higgs boson decays and the SM Higgs boson decay $H \to \tau\tau$. The Top, $Z \to \tau\tau$ and fake-leptons background backgrounds also peak in the same vicinity of the signal peak. The only backgrounds that the BDTs are able to discriminate against are the $Z \to ll$ and to some extent the $WW/WZ/ZZ$ background, that behaves very much like the $Z \to ll$ background in some of the input variables.
This is most likely an effect of the BDTs being overwhelmed by the dominant $Z \to ll$ background. This causes an effective blindness of the BDTs to sufficient small background contributions in their training, as the BDTs become only sensitive to a tiny background contribution if a specific node of the BDT is sufficiently enriched in that background. If a background contribution is too small, a stopping criterion of the BDT training is met before a such node is formed.
A possible solution for this problem would have been to give all backgrounds the same weight during the training, causing most likely a worse performance of the BDT with respect to suppressing the dominant $Z \to ll$ background. Therefore it was decided to pursue the strategy of cutting on the BDT score of BDTs_1 and training with BDTs_2 a second set of BDTs. This is described in the next section. This approach increases the sensitivity of the analysis significantly. It furthermore avoids

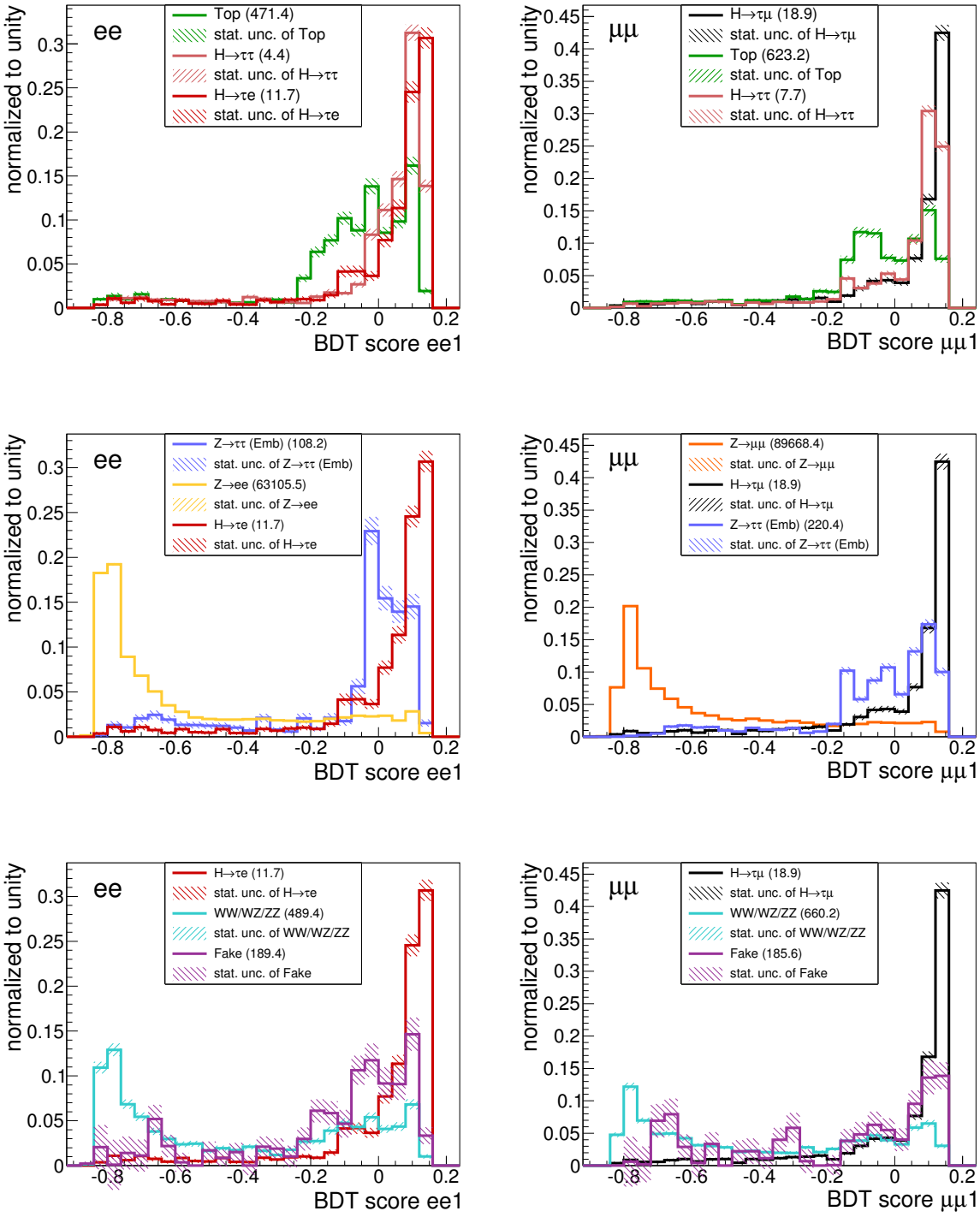


Figure 7.19: The left (right) plots show the shape of signal compared to the shape of multiple background processes for the signal region of the ee ($\mu\mu$) final states for the corresponding BDT score. The contributions of the ggH and VBF production modes of the various Higgs decays have been added up. All shown distributions are normalized to unity and the given uncertainties are only statistical. The number stated in the brackets corresponds to the expected event yield.

implications of the peak structure of some of the backgrounds at a high BDT score, that could limit the fits capability to extract the signal from data, as a background alone, or the sum of some of the backgrounds, could either fake or mask a signal.

7.5.2 Second Set of BDTs

Event Selection

The BDT score from BDTs_1 is now used to split the events from each final state into a signal and $Z \to ll$ control region. The cut SF6 that is used for this splitting is given in table 7.10. It is applied on top of the cuts of the basic VBF selection presented in section 5.3 and the cuts SF1 to SF5 from the previous section. Cut SF6 was manually selected as it suppresses the $Z \to ll$ background

	ee final state		**$\mu\mu$ final state**	
Cut name	ee Signal Region	$Z \to ee$ CR	$\mu\mu$ Signal Region	$Z \to \mu\mu$ CR
SF6	BDT score $ee1 > 0.05$	BDT score $ee1 < 0.05$	BDT score $\mu\mu1 > 0.05$	BDT score $\mu\mu1 < 0.05$

Table 7.10: Additional cut for the training of the second BDTs in the same flavor final states. Events are kept, if the expression given in the table evaluates as true.

significantly, while sparing most of the signal. It was not optimized, as an optimization of this cut is very time consuming, since a new set of BDTs would have to be trained for each cut value.
The cut SF6 introduced a control region for the $Z \to ee$ and $Z \to \mu\mu$ background that is orthogonal to the corresponding signal region. These control regions are now used to determine the normalization of the corresponding $Z \to ll$ background, by scaling this background such that the background prediction matches the observed data. Table 7.11 gives the new normalization of the $Z \to ll$ backgrounds and the used scale factors for the embedding and fake-lepton background estimate. It can be seen that the normalization factors of the $Z \to \mu\mu$ and $Z \to ee$ background estimates are almost identical to the ones calculated previously in the Z-peak region listed in given in table 7.7. The derived normalizations of the $Z \to \mu\mu$ and $Z \to ee$ background is only used as an estimate for the normalization in the training of the BDTs to ensure that they are trained with a approximately correct relative background contributions. These normalizations are free parameters in the fits performed in the statistical evaluation, allowing the fit to find on the best normalization of these backgrounds.

Norm factor	**ee final state**	**$\mu\mu$ final state**
$Z \to ee$ normalization	1.126	n.a
$Z \to \mu\mu$ normalization	n.a.	0.999
Embedding normalization	0.192	0.136
Fake scale factor	0.22 ± 0.03	0.337 ± 0.017

Table 7.11: Scale and normalization factors used in the same flavor final states. The stated normalization factors for the $Z \to ee$ and $Z \to \mu\mu$ background replace the preliminary normalization factors used in the previous section.

The expected event yield for the signal and control regions[9] in the same flavor final states are given

[9] It can be seen that the total background event yield in the $Z \to \mu\mu$ and $Z \to ee$ control regions does not match exactly the observed data as it is expected due to some technical issue. The relative discrepancy is however at the order of 10^{-5} and therefore negligible.

in table 7.12. Less than 5 % of the $Z \to ll$ background of each final state pass the selection of the signal region, while about 65 % of the signal events pass the selection of the signal regions, enhancing the signal to background ratio by approximately a factor of 13 in both final states. The sum of the backgrounds in the signal regions is still dominated by the $Z \to ll$ background, but the the relative contribution from non-$Z \to ll$ backgrounds increased significantly from 2 % to 8 % (2 % to 9 %)in the ee ($\mu\mu$) final state.
The expected event yields of the $\mu\mu$ final state are in good agreement with the observed data, but there is a shortage of events predicted by the background estimates compared to the observed data in the ee final state, that is not covered by the given statistical uncertainties. There are multiple systematic uncertainties in the ee final state, like the jet energy resolution or the energy scale of the electrons, that can change the expected event yield of the $Z \to ee$ background in the signal region by more that 10 % within a $1\,\sigma$ variation. This discrepancy between the observed data and the background estimates in the signal region of the ee final state is therefore non significant.

	ee final state		**$\mu\mu$ final state**	
	ee Signal Region	$Z \to ee$ CR	$\mu\mu$ Signal Region	$Z \to \mu\mu$ CR
$Z \to \mu\mu$	0.0 ± 0.0	0.0 ± 0.0	4193.2 ± 52.4	85228.9 ± 198.6
$Z \to ee$	2874.0 ± 41.8	60537.7 ± 176.5	0.0 ± 0.0	0.0 ± 0.0
top quark	125.0 ± 6.1	346.4 ± 9.7	192.8 ± 7.6	430.4 ± 11.5
WW/WZ/ZZ	56.4 ± 3.7	433.0 ± 9.5	93.1 ± 5.2	567.1 ± 11.1
Fake-leptons	47.4 ± 4.9	141.9 ± 13.7	65.9 ± 7.1	115.0 ± 18.6
$Z \to \tau\tau$ (Emb.)	27.1 ± 2.0	81.1 ± 3.2	83.1 ± 2.6	137.4 ± 3.3
VBF $H \to \tau\tau$	1.6 ± 0.0	0.9 ± 0.0	3.0 ± 0.0	1.2 ± 0.0
ggH $H \to \tau\tau$	0.8 ± 0.1	1.1 ± 0.1	2.0 ± 0.1	1.6 ± 0.1
VBF $H \to \tau\mu$	0.0 ± 0.0	0.0 ± 0.0	7.3 ± 0.1	2.2 ± 0.1
ggH $H \to \tau\mu$	0.0 ± 0.0	0.0 ± 0.0	5.1 ± 0.3	4.3 ± 0.3
VBF $H \to \tau e$	4.5 ± 0.1	1.4 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
ggH $H \to \tau e$	3.0 ± 0.2	2.8 ± 0.2	0.0 ± 0.0	0.0 ± 0.0
Sum $H \to \tau\mu$	$\mathbf{0.0 \pm 0.0}$	$\mathbf{0.0 \pm 0.0}$	$\mathbf{12.4 \pm 0.3}$	$\mathbf{6.5 \pm 0.3}$
Sum $H \to \tau e$	$\mathbf{7.5 \pm 0.2}$	$\mathbf{4.2 \pm 0.2}$	$\mathbf{0.0 \pm 0.0}$	$\mathbf{0.0 \pm 0.0}$
Sum background	$\mathbf{3132.4 \pm 42.8}$	$\mathbf{61542.0 \pm 177.6}$	$\mathbf{4633.1 \pm 53.8}$	$\mathbf{86481.5 \pm 200.1}$
Observed data	$\mathbf{3309.0 \pm 57.5}$	$\mathbf{61544.0 \pm 248.1}$	$\mathbf{4695.0 \pm 68.5}$	$\mathbf{86487.0 \pm 294.1}$

Table 7.12: Expected event yields and observed (obs.) data of the same flavor final states, after splitting the events of each final state in a signal and control region (CR). All given uncertainties are of purely statistical nature.

Table 7.12 shows as well that there is no observable contamination with wrong signal in either final state. In fact, not a single simulated $H \to \tau\mu$ ($H \to \tau e$) signal event passes the selection of the ee ($\mu\mu$) final state. This can be understood since a miss-labeling of the leptons in the same flavor final states does not lead to the appearance of the signal in the wrong final state, as in the different flavor final states. For the appearance of a signal event in the wrong same flavor final state it is, for example for a $H \to \tau\mu$ signal event, necessary that the muon is not detected, the τ decays into an electron and another electron is associated with the event e.g. from a misidentified jet. This is very unlikely and causes that not a single signal event passes the selection of the wrong same flavor final state. Therefore it is not necessary to make any assumptions on the contamination with the wrong signal for the same flavor final states.

BDT Training

The events from the signal region are used to train a new set of BDTs. Table 7.13 lists the variables that were found to contribute significantly to the performance and were therefore used for the BDTs in the analysis. It gives additionally the rank that was assigned to the variable the BDTs. A high rank (small number) correspond to a high usage of this variable by the BDTs.

Variable	**Ranks in*ee* final state**	**Ranks in $\mu\mu$ final state**
ΔR_{ll}	1	2
m_{ll}	2	4
$S(d0_{l_\tau})$	3	1
m_{coll}	4	6
m_{jj}	5	9
E^T_{miss}(hpto)	6	13
$m^T(E^T_{\text{miss}}, l_\tau)$	7	7
x_{l_τ}	8	not considered
$\Delta\eta_{jj}$	9	10
x	not considered	11
E^T_{miss}	10	12
$\Delta\phi(l_\tau, E^T_{\text{miss}})$	11	8
$M_{\text{coll},\tau\tau}$	12	3
$d0_{l_\tau}$	13	5

Table 7.13: Variables and their rank used in the training of the second BDTs in the *ee* and $\mu\mu$ final state.

The number of used variables increased compared to the first set of BDTs. This can be explained by the fact that the events in the signal region behave very signal-like, since the first set of BDTs assigned them a high BDT sore. This means that the differences between signal and background are less pronounced in all variables that were already used in the training of the first set of BDTs. Therefore, the second set of BDTs makes use of additional variables showing a separation between signal and background, even if this separation was not large enough to be considered in the training of the first set of BDTs.
The $d0_{l_\tau}$ and $S(d0_{l_\tau})$ variables were reintroduced to the set of variables, since they seem to be modeled sufficiently well by the background estimates in this part of phase-space. They improve the performance of the BDTs significantly, as these variables show a good separation between signal and background and provide information on l_τ that is independent of the other properties of l_τ that are used in the BDTs.
The BDTs trained for both final states make use of almost the same set of variables, as the $Z \to \mu\mu$ and $Z \to ee$ backgrounds behave very similar and the composition of the other backgrounds is comparable. The distributions of all input variables for the second BDTs in the *ee* final state are shown in figures 7.20, 7.21 and 7.22, while the corresponding distribution of the $\mu\mu$ final state are shown in figures 7.23, 7.24 and 7.25.
The correlations of the input variables is shown in figure 7.26. Signal and background show different correlations between the variables. This can be exploited by the BDTs to discriminate between signal and background.

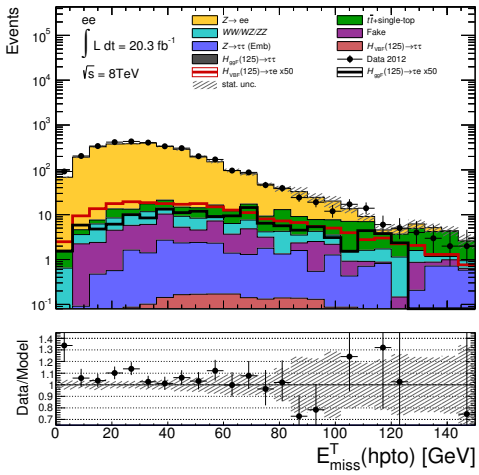


Figure 7.20: Distribution of the input variable E^T_{miss}(hpto) for the training of the second BDTs in the ee final state. The definition of this variable can be found in section 7.2. The plot shows in the upper box the stacked background contributions and the observed data (black dots). Additionally, the $H \to \tau e$ signal that is enhanced by a factor of 50 is shown as a red (black) line for the VBF (ggH) production process. The lower box gives the ratio of the observed data and the sum of all background models. All shown uncertainties correspond only to statistical uncertainties.

Binning

The binning of the BDT score of the same flavor final states is selected analogously to the binning in the different flavor final state according to the strategy discussed in section 7.3.3. Figure 7.27 shows the Asimov significance of the highest bin in the BDT score depending on lower bin edge x_0. The values of x_0 that lie in the red shaded area cannot be selected due to the requirements on the background. It is required that each considered background has a contribution in the highest bin with a relative statistical uncertainty of less that 100 %.
The selected lower value of x_0 for the lower bin edge is shown in green. This is the lower bin edge with the best Asimov significance that fulfills the requirements on the background. The selected x_0 of the $\mu\mu$ final state has the best Asimov significance of all tested values of x_0, while the x_0 with the overall best Asimov significance can not be selected in the ee final states due to the requirements on the backgrounds.
The selected values for x_0 are 0.66 for the $\mu\mu$ final state and 0.57 for the ee final state. This selection of the x_0 values yields an Asimov significance of $S_a = 0.94 \pm 0.09\,\sigma$ for the highest bin of the $\mu\mu$ final state that is significantly larger than the Asimov significance of $S_a = 0.77 \pm 0.07\,\sigma$ archived in the the ee final state. This is due to the fact that it is not allowed to select the x_0 values yielding the highest Asimov significance for the ee final state in contrast to the $\mu\mu$ final state.
Additionally, it was decided to shift the high border of the the highest bin in BDT score to 0.74 and 0.72 for the $\mu\mu$ and ee final state, respectively, as the BDTs of the same flavor final states do not make use of the high range of the BDT score. It was carefully checked, that no data or simulated events were lost by neglecting the high BDT region due to the shift of the upper bin border of the highest bin in the BDT score.
This procedure fixes the bin width of the highest bin of the BDT score. The width of this bin is used for the all bins in the high BDT score region, since they contain a reasonable fraction of the

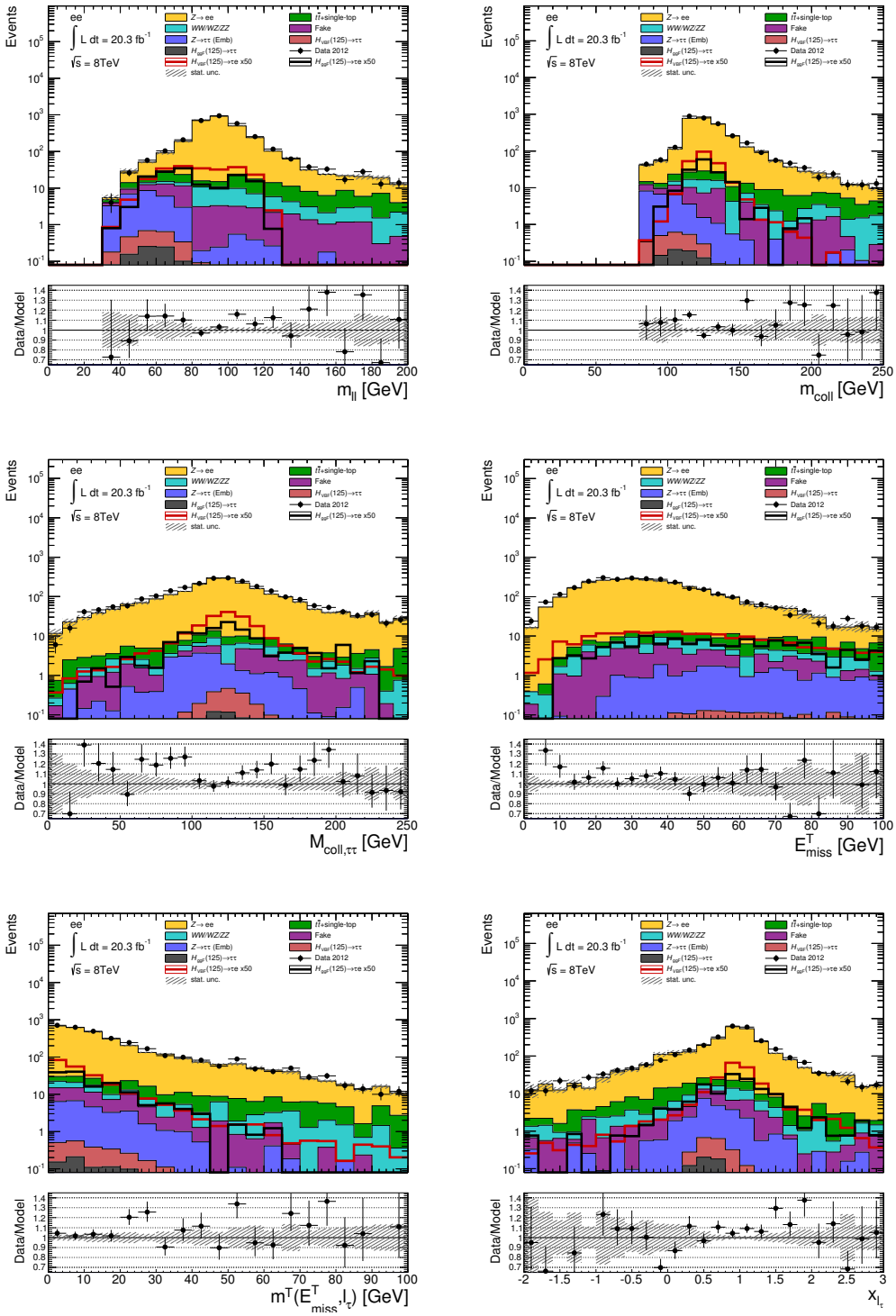


Figure 7.21: Input variables for the training of the second BDTs in the *ee* final state. The definitions of the variables can be found in section 7.2. All plots show in the upper box the stacked background contributions and the observed data (black dots). Additionally, the $H \to \tau e$ signal that is enhanced by a factor of 50 is shown as a red (black) line for the VBF (ggH) production process. The lower box gives the ratio of the observed data and the sum of all background models. All shown uncertainties correspond only to statistical uncertainties.

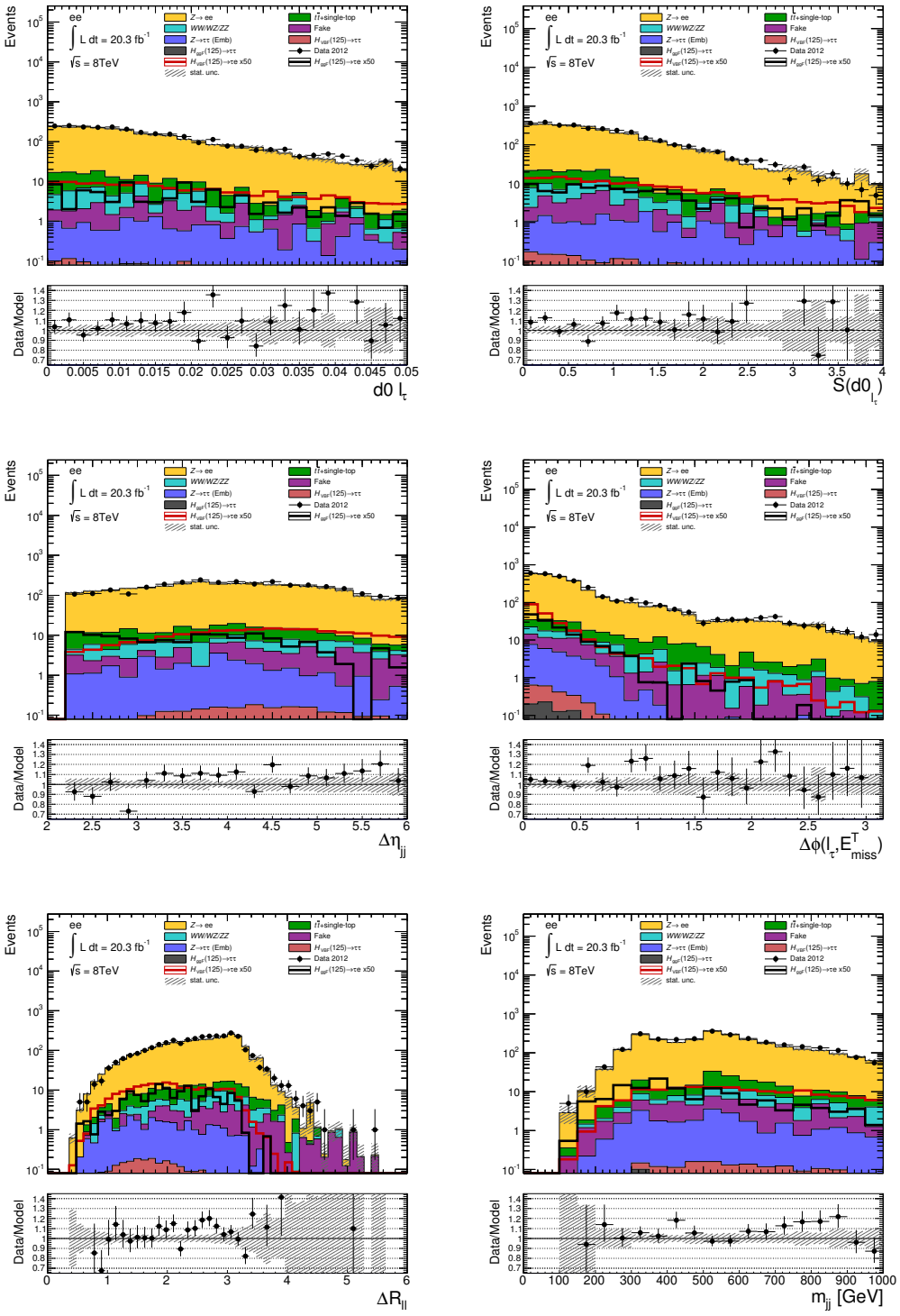


Figure 7.22: Distributions of input variables for the training of the second BDTs in the *ee* final state. The definitions of the variables can be found in section 7.2. All plots show in the upper box the stacked background contributions and the observed data (black dots). Additionally, the $H \rightarrow \tau e$ signal that is enhanced by a factor of 50 is shown as a red (black) line for the VBF (ggH) production process. The lower box gives the ratio of the observed data and the sum of all background models. All shown uncertainties correspond only to statistical uncertainties.

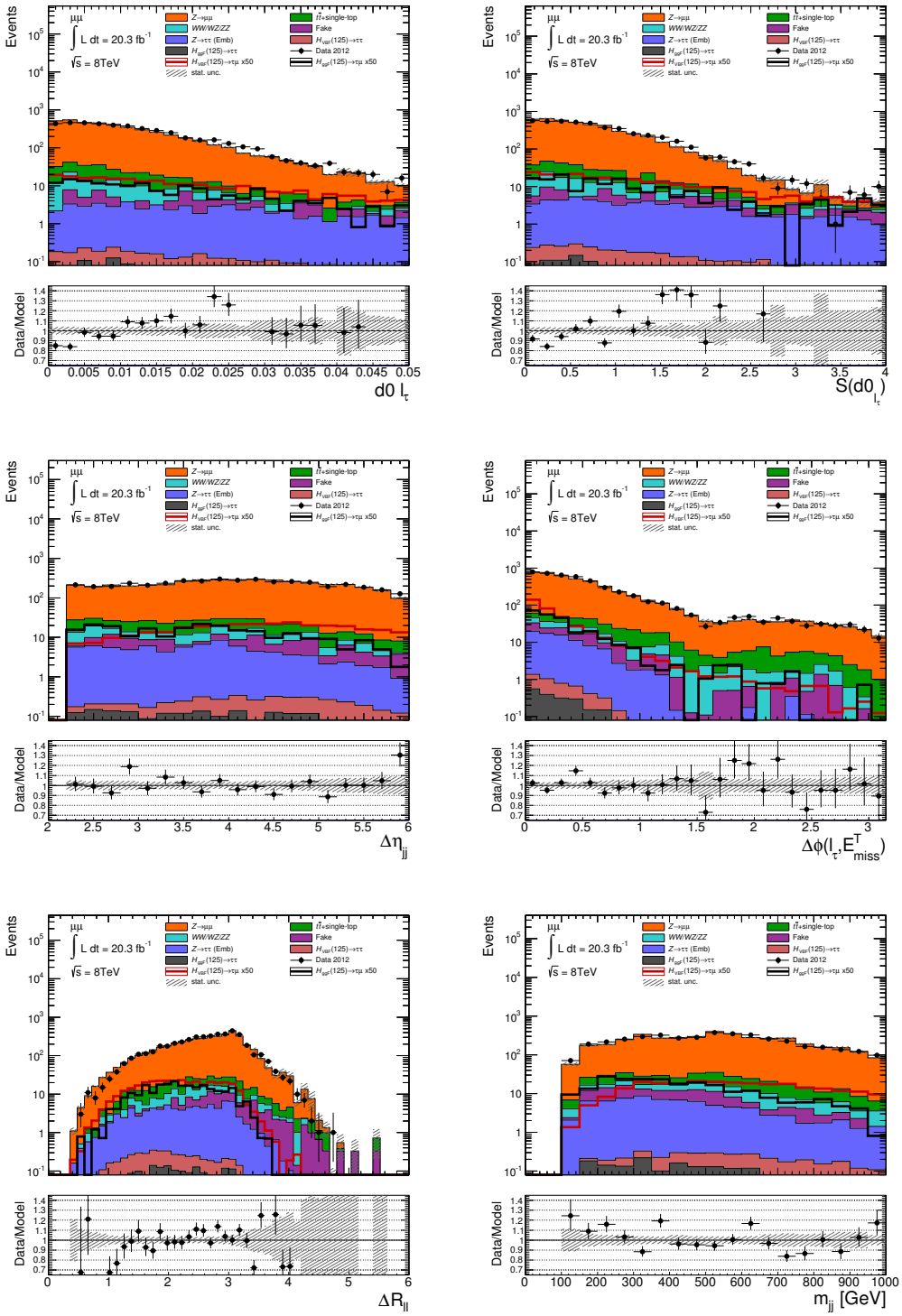


Figure 7.23: Distributions of input variables for the training of the second BDTs in the $\mu\mu$ final state. The definitions of the variables can be found in section 7.2. All plots show in the upper box the stacked background contributions and the observed data (black dots). Additionally, the $H \rightarrow \tau\mu$ signal that is enhanced by a factor of 50 is shown as a red (black) line for the VBF (ggH) production process. The lower box gives the ratio of the observed data and the sum of all background models. All shown uncertainties correspond only to statistical uncertainties.

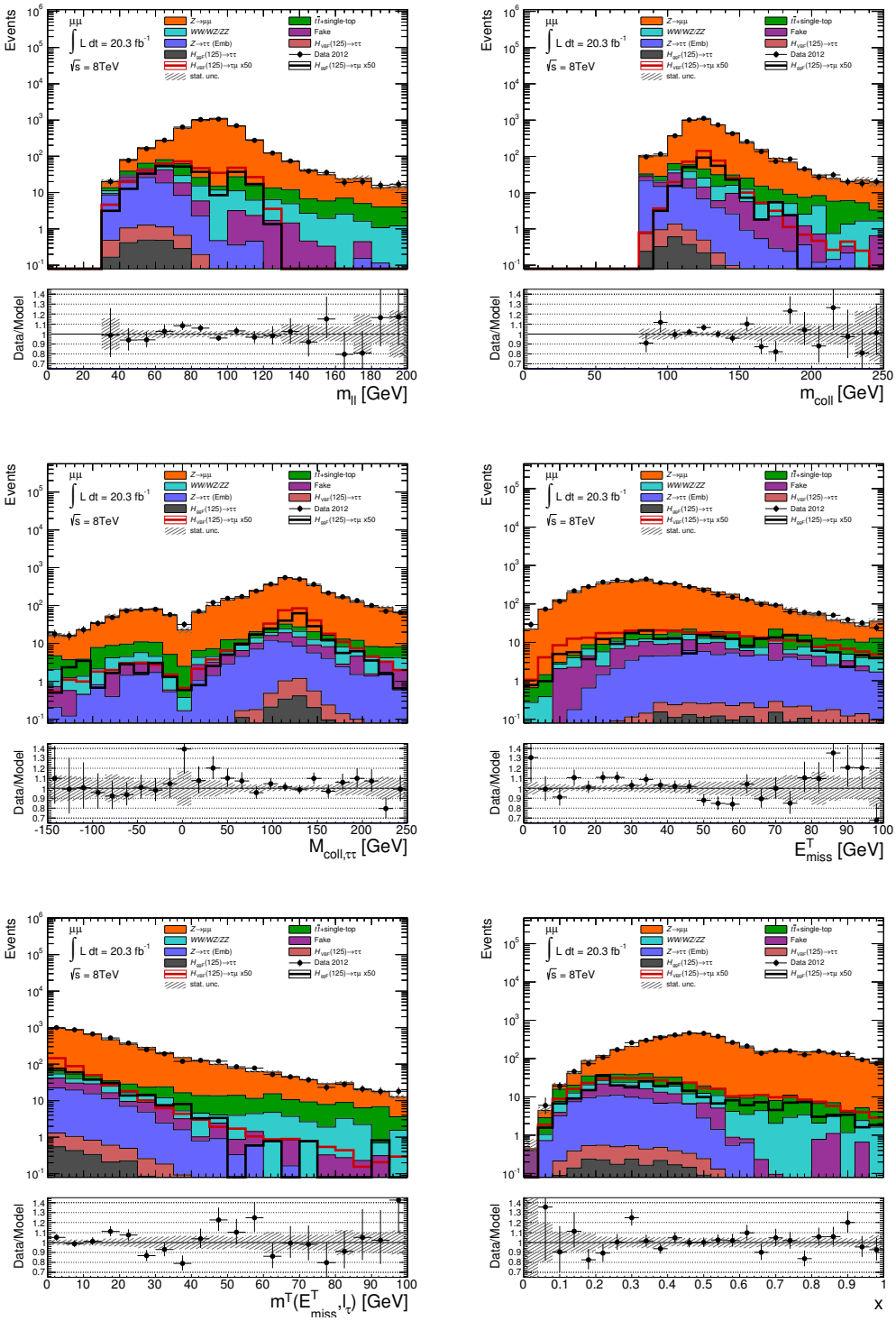


Figure 7.24: Distributions of input variables for the training of the second BDTs in the $\mu\mu$ final state. The definitions of the variables can be found in section 7.2. All plots show in the upper box the stacked background contributions and the observed data (black dots). Additionally, the $H \to \tau\mu$ signal that is enhanced by a factor of 50 is shown as a red (black) line for the VBF (ggH) production process. The lower box gives the ratio of the observed data and the sum of all background models. All shown uncertainties correspond only to statistical uncertainties.

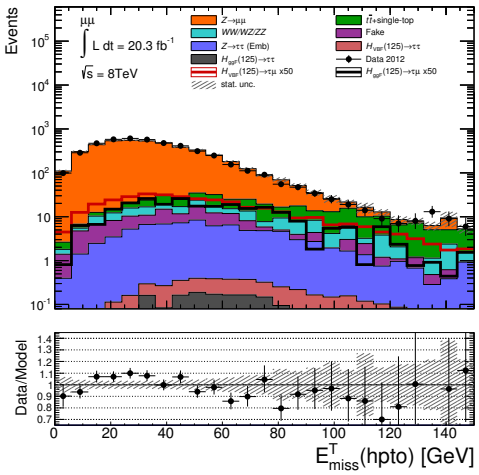
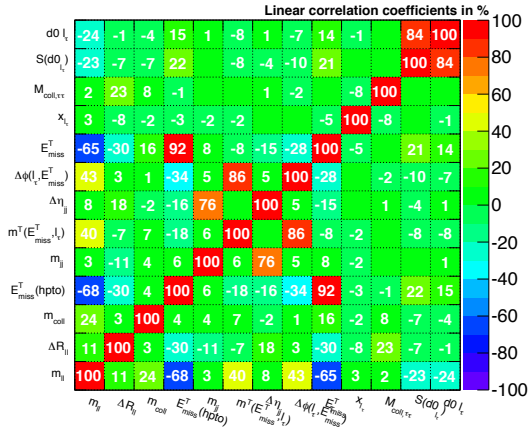


Figure 7.25: Distribution of the input variable E^T_{miss}(hpto) for the training of the second BDTs in the $\mu\mu$ final state. The definition of this variable can be found in section 7.2. The plot shows in the upper box the stacked background contributions and the observed data (black dots). Additionally, the $H \to \tau\mu$ signal that is enhanced by a factor of 50 is shown as a red (black) line for the VBF (ggH) production process. The lower box gives the ratio of the observed data and the sum of all background models. All shown uncertainties correspond only to statistical uncertainties.

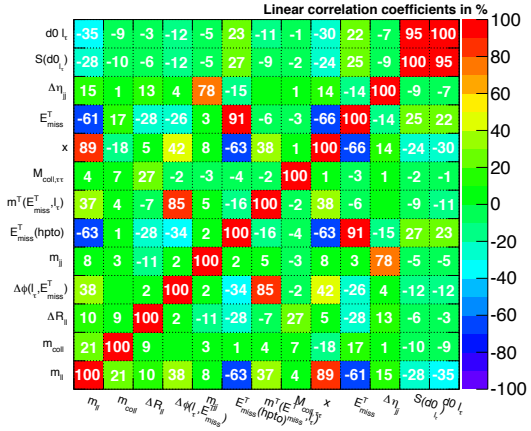
signal and might contribute to the sensitivity of the analysis. The bins in the medium BDT score region are selected such that they have a bigger bin with. This is done to protect them against statistical fluctuations. This does not hurt the sensitivity, as these bins have already a very poor signal to background ratio and a low overall event yield. The low BDT region is combined to one large bin, because a smaller binning is not suitable, as this would yield bins that have no predicted contribution from all backgrounds. Such empty bins would be cumbersome in the fitting procedure, as discussed in section 7.3.3.

BDT Score

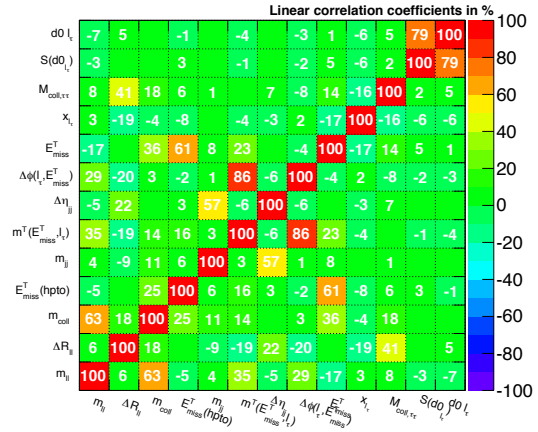
Figure 7.28 shows the distribution of the BDT score of the second set of BDTs in the signal region and the composition of the $Z \to ll$ control regions for both final states. The binning shown in the plots is the final binning used in the statistical evaluation. The binning of the signal regions corresponds to the optimized binning that was introduced in the previous section. All events from the $Z \to ll$ control regions are combined to one bin, since it will only be used by the fit to infer the normalization of the $Z \to ll$ background.
The second set of BDTs is capable to identify the signal better than the first set of BDTs. This is indicated on one hand by the capability of the BDTs to assign higher BDT scores and on the other hand by a significantly improved signal to background ratio in the highest bins of the BDT score. This improved signal to background ratio can bee seen by comparing the distributions of both BDT scores given in figure 7.28 for the second sets of the BDTs and in figure 7.18 for the first set of BDTs. The signal shown in all distributions is enhanced by a factor of 50. The plotted signal of the second sets of BDTs is able to overshoot the plotted background estimates in the highest bins, while this does not happen for the first sets of BDTs. The assignment of higher BDT scores means that more of the individual decision trees are capable to identify signal-like events as signal,



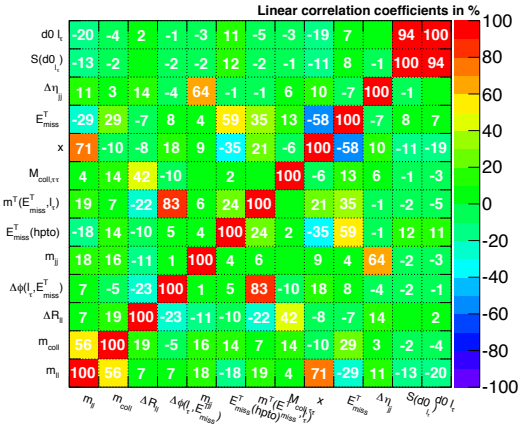
(a) VBF $H \to \tau e$ in the signal region of the ee final state



(b) VBF $H \to \tau\mu$ in signal region of the $\mu\mu$ final state



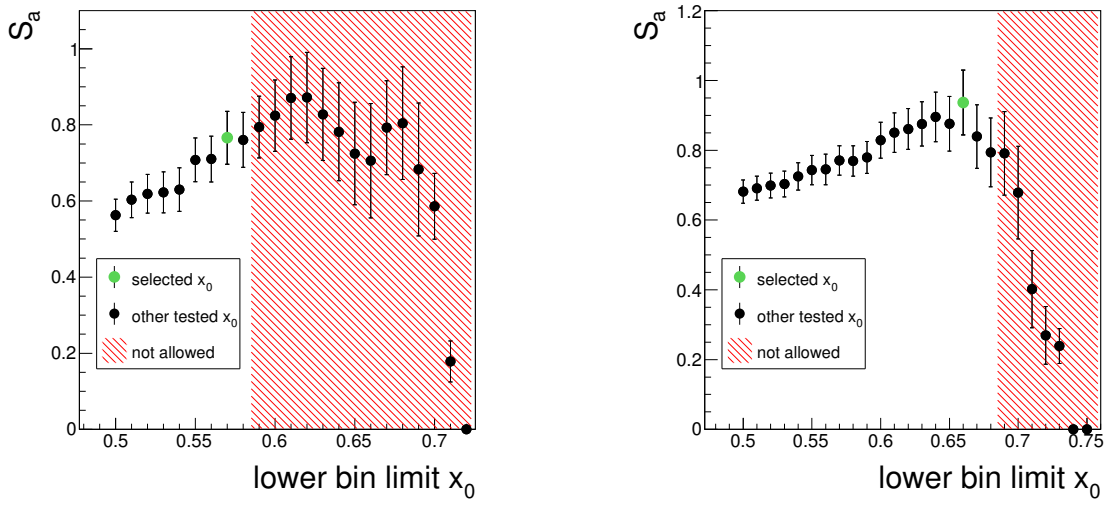
(c) weighted sum of all background processes in the signal regon of the ee final state



(d) weighted sum of all background processes in the signal region of the $\mu\mu$ final state

Figure 7.26: The linear correlation coefficient of the input variables of the second BDTs of the ee ($\mu\mu$) final state on the left (right) hand side. The plots on top are for the signal used in the BDT training, while the plots on bottom correspond to the weighted sum of all background samples used in the BDT training.

and is therefore an indication of an improved performance of the BDTs.
The first set of BDTs showed the cumbersome behavior of not being able to separate between signal and all considered backgrounds, as some of the backgrounds showed a peak-structure in the BDT score that is very similar to the one of the signal. The separation between the peaks of signal and backgrounds in the BDT score has improved significantly compared to the first set of BDTs. This can be seen in figure 7.29 that compares the shape of the signal with all considered backgrounds for the second set o BDTs. None of the backgrounds peaks in the direct vicinity of the peak of the signal in the highest BDT score.
The $Z \to ee$ and $Z \to \mu\mu$ control region, that are shown in figure 7.28 (b) and (d), are very pure in the $Z \to ee$ and $Z \to \mu\mu$ background, respectively, justifying their usage in the derivation of the normalization of the corresponding $Z \to ll$ background.

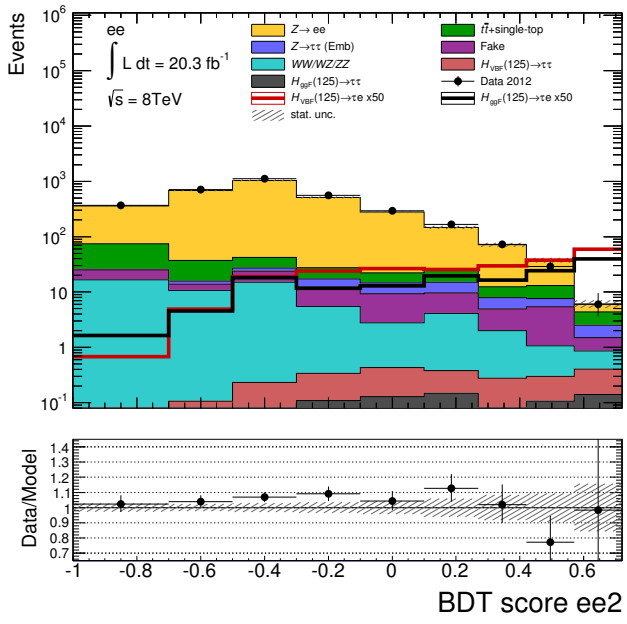


(a) ee final state

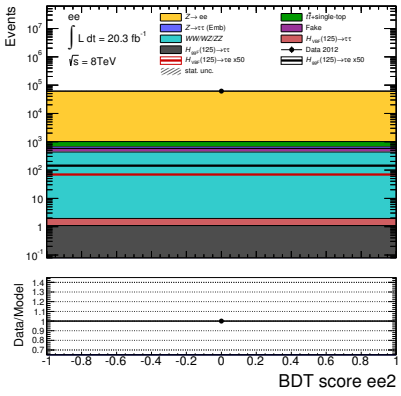
(b) $\mu\mu$ final state

Figure 7.27: Asimov significance S_a of the highest bin of the BDT score that contains all event in the interval $[x_0, 1]$ depending on the lower bin bound x_0. The left (right) plot shows the ee ($\mu\mu$) final state. The Asimov significance corresponding to the selected x_0 is marked in green, while the Asimov significance of all other tested x_0 are shown in black. The selection of a x_0 lying in the red area is not allowed by the requirement on the background predictions. The shown uncertainties are inferred from the statistical uncertainties of the signal and background predictions.

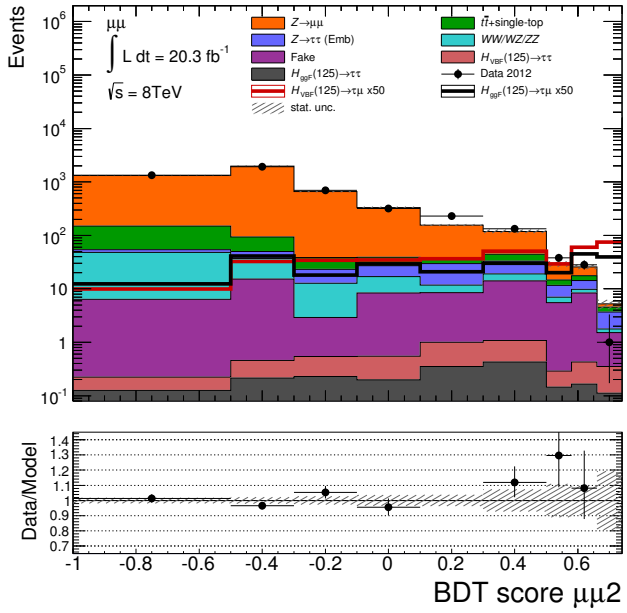
Unfortunately it can bee seen that the BDT scores of the second BDT are not modeled very well by the background estimates. The previously discussed overall too small background prediction of the signal region of the ee final state causes a tension between data and the background estimates, as can be seen in figure 7.28 (a). Almost all bins show a shortage of the background estimate compared to data. Nevertheless, the shape of the BDT score of the signal region of the ee final state is modeled sufficiently. The BDT score of the $\mu\mu$ final state is shown in figure 7.28 (c). It can be seen that the modeling of the shape of the BDT score has some issues in this final state, while the background estimates predict the overall event yield well. There are two bins, the highest bin and the bin in the vicinity of a BDT score of 0.2, with a significant gap between the background predictions and the observed data.
It was unfortunately not possible to track down the sources of the miss-modeling and correct them within the timescale of this master's thesis. Therefore the distributions of the second BDT score were used in the statistical evaluation, although they show the previously discussed tensions between data and the background prediction.
The presence of a large signal would be visible in the highest bin of the BDT score, as it provides a reasonable signal to background ratio. The highest bin of the BDT score shows a shortage of observed data in the $\mu\mu$ final state and a good agreement between observed data and the background estimates in the ee final states. This means that there is no obvious footprint of a signal in the same flavor final states.



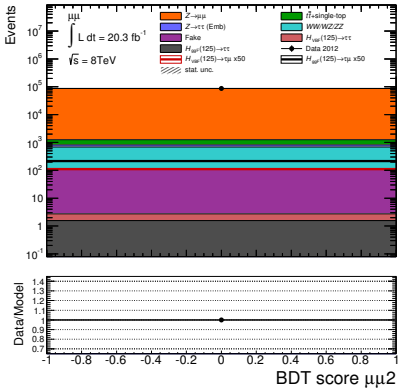
(a)



(b)



(c)



(d)

Figure 7.28: Distribution of the BDT score in the signal region and composition of the $Z \to ll$ control regions of both same flavor final states. Figure (a) and (b) correspond to the *ee* final state, while Figure (c) and (d) correspond to the $\mu\mu$ final state. The shown signal is enhanced by a factor of 50.

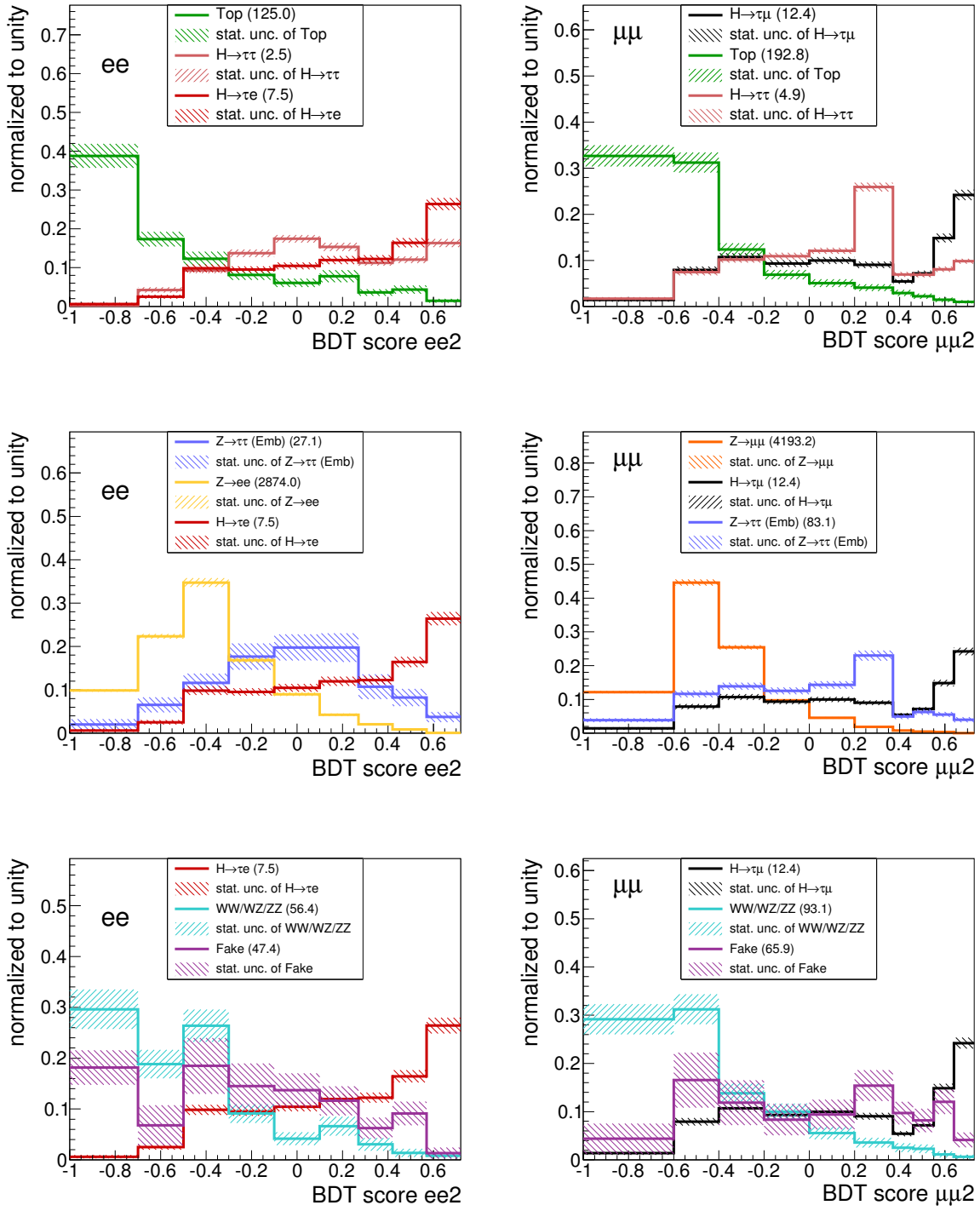


Figure 7.29: The left (right) plots show the shape of the signal compared to the shape of multiple background processes for the signal region of the ee ($\mu\mu$) final states for the corresponding BDT score. The contributions of the ggH and VBF production modes of the various Higgs decays have been added up. All shown distributions are normalized to unity and the given uncertainties are purely statistical. The number stated in the brackets correspond to the expected event yield.

Chapter 8

Systematic Uncertainties

This chapter gives an overview of the considered systematic uncertainties and their treatment. Systematic uncertainties arise typically if a quantity is used in the analysis that is not known precisely. Such a quantity can be for example the energy scale of the electrons. These quantities are measured carefully and their best estimated value, denoted by the nominal value, is used in the analysis as an estimator for them. The nominal value is nevertheless not necessarily the correct value and hence introducing a bias to the analysis. This possible bias of the analysis is usually estimated by changing the considered quantity from its nominal value to its nominal value plus (minus) its estimated upper (lower) uncertainty and inferring the effects of this on the distribution of the final discriminant of the analysis in terms of the variation of the event yield and the variation of the shape. These effects are then considered as additional uncertainties in the fit model.

8.1 Considered Systematic Uncertainties

This analysis uses almost the same set of systematic uncertainties as the $H \to \tau\tau$ analysis of ATLAS. This thesis describes therefore only uncertainties in detail that are specific to the presented analysis. A more detailed description of the systematics that were also used in the $H \to \tau\tau$ analysis can be found in reference [37]. The considered systematic uncertainties can be divided into three groups: experimental uncertainties, background modeling uncertainties and theoretical uncertainties of the Higgs boson samples.

8.1.1 Experimental Uncertainties

This group includes the important experimental uncertainties originating from uncertainties on the efficiencies for triggering, object identification and reconstruction, uncertainties on the energy scale and resolution of objects and the uncertainty on the luminosity. This group of uncertainties is not discussed in detail, since the presented analysis uses exactly the same experimental uncertainties as the SM $H \to \tau\tau$ [37] analysis of the ATLAS collaboration.

8.1.2 Background Modeling Uncertainties

The following list gives the uncertainties that affect only specific backgrounds.

- $Z \to \tau\tau$ **(Emb.)**: This analysis applies the same systematic uncertainties on the embedding background estimate as the SM $H \to \tau\tau$ analysis of the ATLAS collaboration.

- **Fake-leptons**: Two uncertainties are considered for the estimation of the fake-lepton background. The first uncertainty, denoted by *FakeHistoSys*, is calculated by repeating the fake estimate in a same-sign control region with the same event selection as in the main analysis, but asking for two leptons of the same charge. The template for the fake-lepton background that was extracted in the same-sign control region is then scaled to the expected event yield of the signal region times the difference of the fake-scale-factors of the signal and control region normalized to the fake-scale-factor of the signal region. This modified prediction of the fake background is then used as an estimate for the uncertainties on the derivation of the fake-lepton estimate. Table 8.1 gives the uncertainties on the event yield of each final state due to this uncertainties. They range between 7.2 % and 43.4 %.
  Since the fake-lepton estimate is known to be not modeled precisely, a second uncertainty of ±20 % is applied on the event yield of this background estimate, to make sure that a potential small mismodeling of this background is covered by the applied uncertainties, as in the SM $H \to \tau\tau$ analysis of the ATLAS collaboration.

Final state	Effect on event yield
$\mu\mu$	±14.8 %
ee	±43.4 %
μe	±7.9 %
$e\mu$	±7.2 %

Table 8.1: Effects of the uncertainty *FakeHistoSys* on the event yield of the fake-lepton background in the four final states.

- **Top quark background:** It was not possible to realize a data driven normalization of the top quark background in the same flavor final state, due to some issues in the modeling of the $Z \to \mu\mu$ and $Z \to ee$ background in that specific phase space (for more details on this see section 7.5.1). Therefore additional uncertainties on the event yield of the Top background are applied in the same flavor final states: one for the uncertainty of the QCD scale (+2.53 %; −3.40 %) and one for for the uncertainty on the PDFs and α_s (±4.63 %). These uncertainties were estimated from the theoretical uncertainties on the top quark production cross section from reference [114]. They do however not cover any effects on the shape of the Top backgrounds.

- **$Z \to ee$ in the ee final state:** It was necessary to introduce scale factors for the $Z \to ee$ background in the derivation of the scale factor for the fake-leptons background of the ee final state, as discussed in section 4.2.3. An uncertainty on this is estimated by rerunning the analysis without applying the scale factors. The obtained distribution of the BDT score is then used as an estimate for the effects of potential systematic uncertainties caused by the reweighing procedure. It has an effect of 0.8 % (0.03 %) on the event yield of the $Z \to ee$ in the ee signal region ($Z \to ee$ control region).

8.1.3 Theoretical Uncertainties of the Higgs Boson Samples

The timescale of this master's thesis did not allow to calculate the theoretical uncertainties on the Higgs boson samples in detail. They are nevertheless important, since they can have a large impact on the event yield of the Higgs boson MC samples. Therefore, it was decided to use the effect of

these systematic uncertainties on the event yield of the fully leptonic final state of the SM $H \to \tau\tau$ analysis of the ATLAS collaboration [37] as an estimate. This procedure does obviously not allow to consider effects of these uncertainties on the shape of the BDT score distributions.

Production mode	Name	Effect on event yield	
		$-1\,\sigma$	$+1\,\sigma$
ggH	pdf_Higgs_gg_ACCEPT	−10.3 %	+5.5 %
ggH	pdf_Higgs_gg	−6.9 %	+7.7 %
ggH	QCDscale_ggH	−21.0 %	+26.0 %
ggH	ATLAS_UE_gg	−4.0 %	+4.0 %
VBF	pdf_Higgs_qq_ACCEPT	−2.2 %	+0.9 %
VBF	pdf_Higgs_qq	−2.8 %	+2.6 %
VBF	QCDscale_qqH	−2.5 %	+2.5 %
VBF	ATLAS_UE_qq	−1.0 %	+1.0 %
VBF	NLO_EW_Higgs	−2.0 %	+2.0 %

Table 8.2: Effects of the theoretical uncertainties of the Higgs samples in the fully leptonic final states of the $H \to \tau\tau$ analysis of the ATLAS collaboration on the expected event yield. They are used as an estimation for these uncertainties in the presented analysis

The considered uncertainties and their effects on the event yield are listed in table 8.2. They correspond to the theoretical uncertainties in the parton distribution functions (pdf), the uncertainties in the molding of the underlying events (UE), the QCD scale and the uncertainties on the NLO electroweak corrections on the production of a Higgs boson via the VBF production mode. These uncertainties are applied on the LFV signal and the SM Higgs boson background samples, but the uncertainties are assumed to be uncorrelated between these two sets of samples in order to have a more conservative estimate.

8.2 Treatment of Systematic Uncertainties

This section summarizes the treatment of systematic uncertainties. At first, the used methods are introduced, followed by a discussion of the actual treatment of systematic uncertainties.
The systematic uncertainties are handled mostly at the level of histograms of the final discriminant in the binning used in the statistical evaluation. Usually, there is one nominal histogram inferred from the analysis using all nominal values of the error-prune quantities and two histograms for each systematic uncertainties corresponding to the analyses carried out with the value of the quantity linked to the systematic uncertainty being set to the nominal value plus (minus) its upper (lower) uncertainty denoted by up-type (down-type) histogram.

8.2.1 Symmetrization of Missing Histograms

Not all estimates of the systematic uncertainties provide information on both, the upper and lower uncertainty, hence either the up-type, or down-type histogram is not available for these uncertainties. This lack of information on the effects of the upper or lower uncertainty is artificially caused by limitations on the estimates of these uncertainties and therefore not scientifically motivated. Due to this, it is necessary to find a reasonable way to estimate the missing information on the effects of the upper or lower uncertainty on the analysis. This analysis deals with this issue by creating the

missing histogram by symmetrizing the available histogram with respect to the nominal histogram. The algorithm needs two input histograms, the nominal histogram, and the available up- or down-type histogram. For notational convenience it is assumed in the following that the available histogram is the up-type histogram. First, both input histograms are normalized to unity. The content of the missing down-type histogram, is calculated bin-wise. A factor f_i s calculates for each bin i by

$$f_i = \frac{\text{content of bin } i \text{ of up-type histogram}}{\text{content of bin } i \text{ of nominal histogram}}, \tag{8.1}$$

that is capped at 2 meaning that values larger than 2 of f_i are set to 2. This is necessary to limit the bin content of the created histogram to physical meaningful values. The content of bin i and its uncertainty of the down-type histogram is than calculated by scaling the bin content and the uncertainty of bin i of the nominal histogram by the factor $(2 - f_i)$. So far, this procedure symmetrized the effects on the shape of the final histogram. The down-type histogram is now scaled such that its integral shows the same absolute difference to the nominal histogram, as the up-type histogram, but in the opposite direction. Figure 8.1 gives an example of a histogram that was created by the symmetrization process.

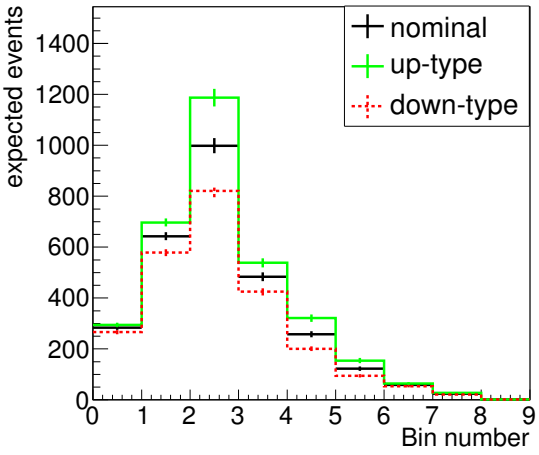


Figure 8.1: Example of the result of the symmetrization process. The down-type histogram, shown in red, was created by the symmetrization process. The corresponding nominal and up-type histograms are given in black and green, respectively.

8.2.2 Pruning of Systematic Uncertainties

The influence of a systematic uncertainty on the shape of the distribution of the final discriminant is tested with a so called pruning procedure. This pruning procedure is done for each of the considered systematic uncertainties separately using the nominal, up- and down-type histograms. In the first step, these histograms are normalized to unity to eliminate effects on the total event yield. Then two χ^2 test are executed: one between the up-type and the nominal histogram and one between the down-type histogram. A systematic uncertainty is considered to not have an effect on the shape of the distribution of the final discriminant if both test yield a χ^2 square probability of more than 95 %. In this case it is said that the systematic is pruned meaning that its effects on the shape of the final discriminant will not be considered in the fit model, as they were found to be insignificant.

8.2.3 Applied Treatment of Systematic Uncertainties

Each systematic uncertainty is tested for each considered process (background and signal estimates) for its impact on the analysis. This testing procedure decides how, or whether at all, a systematic uncertainty is introduced to the fit model for the given process. The whole treatment is visualized in figure 8.2.

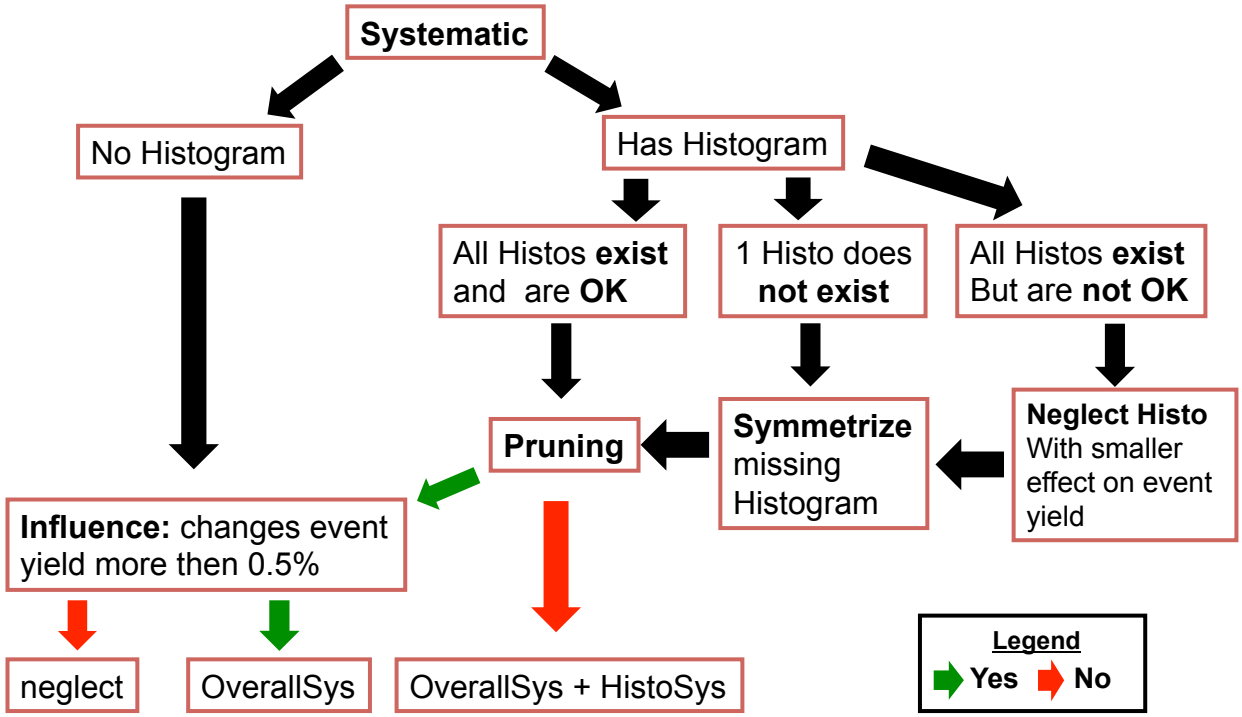


Figure 8.2: Illustration of the treatment of systematic uncertainties in this analysis. *Histo* is used as an abbreviation for histograms.

First it is checked if any histogram is available for this systematic uncertainties. Most systematic uncertainties have associated histograms, except the ones that affect by definition only the event yield and not the shape of the final discriminant. Such uncertainties are within this analysis for example the uncertainty on the luminosity or the applied estimates of the theoretical uncertainties on the Higgs samples. These systematics are checked for influence on the total event yield. They are introduced as an *OverallSys* (Systematic on the overall event yield) to the fit model, if they have an effect on the event yield of more than 0.5 %, and neglected if they do not.
The treatment of systematics that come with at least one histograms is shown in the right branch of figure 8.2. First it is checked if the up- and down-type histograms are available and if the histograms are *OK*. The histograms are considered to be *OK* if the up- and down-type histograms change the event yield in a different direction[1]. If the histograms were found to be not *OK*, the histogram (up- or down-type) with the smaller effect on the yield gets removed from the procedure an the systematic uncertainty is treated from here on in the same way as systematic uncertainties that have only one of the up- or down-type histograms. This is done since it might cause trouble in the fitting procedure if both the upward and downward variation of the systematic causes the

[1]This means that one of the up- or down-type histograms decreases the event yield, while the other is increasing the event yield.

event yield to be shifted in the same direction and it is anyway the more conservative approach.
All systematic uncertainties that have only one of the up- and down-type histograms are given to the symmetrization algorithm to create the missing histogram and are treated from here in the same way as the histograms that had both histograms from the beginning and were found to be OK.
The histograms of these systematic uncertainties are tested by the pruning procedure, described above, if their effect on the shape of the distribution of the final discriminant can be pruned away. If this is the case, they are tested for their influence on the event yield and either introduced as an *OverallSys* or neglected depending if they have an effect on the event yield of more that 0.5 % or not. If their effect on the shape of the final discriminant is not pruned away, they are introduced as an *OverallSys* and a *HistoSys* (Systematic on the shape of the distribution).
The main idea behind this whole process of neglecting the effects of the systematic uncertainties, if they were found to be insignificant, is to keep the fit model as simply as possible. This is desirable as it helps to prevent potential numerical issues and reduces the computational effort in the statistical evaluation.

Chapter 9

Statistical Analysis and Results

This chapter presents the statistical evaluation of the distributions of the final discriminant, the BDT score, that were derived in chapter 7. Fits are used to extract information on the possible present LFV signal from the observed data, by adjusting the distributions of the BDT score of all considered signal and background processes to the observed data. The contributions of the $Z \to ee$ and $Z \to \mu\mu$ background were summed in the same flavor final state due to their small event yield. The sum of these two backgrounds is denoted as $Z \to ll$.
The first section describes briefly the used fit model. It is followed by a section discussing the results of the fitting procedure in each final state. A combined fit for the $H \to \tau\mu$ ($H \to \tau e$) signal is presented in the third section using a simultaneous fit in the μe and $\mu\mu$ ($e\mu$ and ee) final state. The last section of this chapter summarizes the obtained results.

9.1 Fit Model

Extended maximum-likelihood [115] fits are performed to extract information on the signal strength of the LVF Higgs boson decays from the distribution of the BDT scores in the various signal regions and the event yields in the background control regions. The statistical model used in the fits performed in this analysis are built using the tool *HistFactory* [116]. It is used to create the likelihood function that is needed as an input for the fits and the calculation of the upper limits on the branching ratio of the signal. The underlying statistical model and the exact form of the likelihood function is described in detail in reference [116].
The derived upper limits on the branching ratio of the LFV Higgs boson decays limits are calculated using the CL_s modified frequentist formalism [117]. All presented limits are calculated at 95 % confidence level.

9.1.1 Systematic Uncertainties in the Fit Model

Systematic uncertainties are included into the fit model as constrained nuisance parameters. This means that the fit is allowed to pull these parameters from their nominal values, but the likelihood is modified directly by this pull, as it contains a Gaussian term for each of the constrained nuisance parameters.
HistFactory offers two different ways to include systematic uncertainties, *OverallSys* and *HistoSys*. A systematic uncertainty introduced as *OverallSys* considers only the effect of the systematic uncertainty on the event yield of the process and does not consider any effects on the shape of the final discriminant. In contrast to this, a systematic uncertainty introduced as a *HistoSys* considers only

the effects of the systematic uncertainty on the shape of the distribution of the final discriminant. Whether the effect on the event yield and the shape of the distribution or only the event on the event yield is considered in the fit is explained in section 8.2.
Systematic uncertainties are considered to be fully correlated between the signal and control region in each individual final state. The combined fits performed simultaneously in two final states treat most of the systematics as fully correlated. Only the systematics of the fake-lepton background estimate are assumed to be uncorrelated between the final states in those combined fits.

9.1.2 Treatment of Statistical Uncertainties of the Considered Processes

The statistical uncertainties are fully considered in this analysis, since the statistic of most MC samples is relatively small in the selected phase space region of the VBF topology. *HistFactory* treats them as constrained bin-by-bin scale factors that are shared by all considered processes. For more details on this see reference [116].

9.1.3 Free Parameters of the Fit

The fit model includes a handful of free parameters. The fit is allowed to pull these parameter to any value, within their given range, without modifying the likelihood directly. This is possible, as the likelihood does not contain any terms restricting the value of these free parameters directly.
This analysis uses free parameters for two purposes in its fit model. They are used for the parametrization of the signal strength μ and for the normalization of the dominant background contribution in each final state. Table 9.1 gives an overview of the free parameters in the fit model sorted by the final states.

Final state	Name	Range
$\boldsymbol{ee}$	μ $(H \to \tau e)$	$[-10, 10]$
	norm_Zee	$[0.5, 1.5]$
$\boldsymbol{e\mu}$	μ $(H \to \tau e)$	$[-10, 10]$
	norm_TopEMu	$[0.5, 1.5]$
$\boldsymbol{\mu\mu}$	μ $(H \to \tau \mu)$	$[-10, 10]$
	norm_Zmumu	$[0.5, 1.5]$
$\boldsymbol{e\mu}$	μ $(H \to \tau \mu)$	$[-10, 10]$
	norm_TopEMu	$[0.5, 1.5]$

Table 9.1: Free parameters of the fit model sorted by the final states. A value of the signal strength of $\mu = 1$ corresponds to a branching ratio of the corresponding decay of 1 %.

It can be seen that the signal strength μ is allowed to take negative values. It is necessary to introduce these negative values of the signal strength, although they are not physically motivated, as limiting μ to positive values would introduce a bias to the fit model and is therefore not acceptable. This is especially an issue in combined fits with multiple signal regions, where a fit limited to $\mu > 0$ is only able to absorb upward fluctuations in the observed data in a given set of signal regions in the fitted μ, while it is not able to absorb downward fluctuations in another set of signal regions into the fitted μ, introducing obviously a bias.
The normalization of the main background in each final state is inferred from data by keeping it a free parameter in the fit. Chapter 7 introduced a control region for each of these backgrounds to constrain the free floating normalization constants. The normalization factors of the background

are only applied to those final states that include a control region for this specific background, otherwise the backgrounds are normalized according to their estimated event yields.

9.1.4 Asimov Dataset

The so-called Asimov dataset is a toy dataset built from the background and signal estimates, hence does not contain any observed data. It is constructed by adding all pre-fit background estimates and the pre-fit signal expectations that are scaled to a specific input signal strength μ. A $\mu = 1$ corresponds in this analysis to a branching ratio of the considered decay ($H \to \tau\mu$ or $H \to \tau e$) of 1 %. This analysis uses Asimov datasets with an input of $\mu = 0, 1, 5$, to test the performance of the fit. The fit's capability to retrieve the input μ is tested and the observed significance on the Asimov dataset is used as an estimation for the significance for a given μ. The Asimov dataset with $\mu = 0$ is used additionally to extract the expected upper limit on the branching ratio.

9.2 Expected Sensitivity and Results in Individual Final States

This section presents the fits performed separately on each of the four final states. This provides two independent measurements of both possible signals, as each fit in the μe and $\mu\mu$ final states extracts information on the possible $H \to \tau\mu$ signal, while the individual fits in the $e\mu$ and ee extract information on the possible $H \to \tau e$ signal. All final states are treated analogously.

9.2.1 Fit for $H \to \tau\mu$ in the μe Final State

Expected Sensitivity

The performance of the fit is tested using Asimov datasets with different input values for the signal strength μ. A value of $\mu = 1$ corresponds to a branching ratio of 1 % of the decay $H \to \tau\mu$. Table 9.2 gives the tested input values of the signal strength μ, the best fit value $\hat{\mu}$ of μ and the

μ in Asimov data	$\hat{\mu}$	expected significance
0	0.00 ± 0.71	-0.01
1	1.00 ± 0.87	1.30
5	5.00 ± 1.45	5.30

Table 9.2: Performance of the fit in the μe final state on Asimov data with a different input value of the signal strength μ.

expected significance. It shows first that the fit is able to recover the input value of μ well, without a bias. Second, the expected significance shows that the fit has sensitivity for the tested input values μ that are allowed by the indirect limits on the $H \to \tau\mu$. A signal strength of $\mu = 5$ yields for example an expected significance of more that $5\,\sigma$.
Table 9.3 gives the expected upper limit and its uncertainties for the branching ratio of the $H \to \tau\mu$ decay at 95 % confidence level. The limit and its uncertainties were calculated using Asimov data with $\mu = 0$. The expected limit of the fit in this single final state is already better than the indirect limits discussed in section 2.2.2.

$+2\,\sigma$	4.0 %
$+1\,\sigma$	2.6 %
Median	**1.7** %
$-1\,\sigma$	1.2 %
$-2\,\sigma$	0.8 %

Table 9.3: The expected upper limit and its uncertainties on the branching ratio for the decay $H \to \tau\mu$ at a confidence level of 95 % derived from the fit in the μe final state.

Results in Data

No hints for the LFV decay $H \to \tau\mu$ are observed. Table 9.4 summarizes the information on the $H \to \tau\mu$ decay that was extracted by the fit. Applying a fit to the observed data yields a best

$\hat{\mu}$	observed significance	observed limit BR($H \to \tau\mu$)
0.9 ± 0.8	$1.2\,\sigma$	2.5 %

Table 9.4: Best fit value $\hat{\mu}$ for the signal strength μ of the $H \to \tau\mu$ signal, the observed significance and the observed upper limit on the branching ratio (BR) of the decay $H \to \tau\mu$ at 95 % confidence level for the fit in the μe final state.

Process/Region	μe Signal Region			μe Top Quark CR
	Total	2nd last bin	Last bin	Total
Top quark	660 ± 40	6.6 ± 1.8	6.3 ± 0.9	4150 ± 70
$Z \to \tau\tau$ (Emb)	265 ± 15	3.2 ± 0.4	2.03 ± 0.25	68 ± 5
WW/WZ/ZZ	158 ± 12	1.07 ± 0.22	1.10 ± 0.26	20.3 ± 3.3
Fake-leptons	91 ± 17	1.9 ± 0.8	0.14 ± 0.31	38 ± 7
$Z \to ll$	25 ± 4	0.22 ± 0.08	0.160 ± 0.033	2.7 ± 0.5
VBF $H \to \tau\tau$	3.02 ± 0.22	0.26 ± 0.04	0.38 ± 0.06	0.38 ± 0.04
ggH $H \to \tau\tau$	2.7 ± 0.8	0.16 ± 0.05	0.20 ± 0.06	0.44 ± 0.13
VBF $H \to \tau\mu$	5 ± 4	0.7 ± 0.6	2.9 ± 2.6	0.6 ± 0.5
ggH $H \to \tau\mu$	4 ± 4	0.6 ± 0.6	1.2 ± 1.1	0.5 ± 0.5
Total Bkd.	1205 ± 32	13.5 ± 2.4	10.3 ± 3.5	4280 ± 70
Data	1212	12	16	4283

Table 9.5: Expected event yield and data in the μe signal region and in the μe top quark control region (CR) after the fit in the μe final state using best fit-values for all parameters. The composition of the last and 2nd last bin of the BDT score (two highest bins in the BDT score) of the μe signal region is given additionally. The uncertainties of the individual processes are only systematic uncertainties, while the given uncertainty on the total background (Bkd.) correspond to systematic and statistical uncertainties.

fit value $\hat{\mu}$ of the signal strength of $\hat{\mu} = 0.9 \pm 0.8$ which is in good agreement with the SM only hypothesis corresponding to $\mu = 0$. This is emphasized as well by the observed significance of $1.2\,\sigma$. The mild observed excess of events in the highest bin of the BDT score in the signal region of this final state causes nevertheless the observed limit of 2.5 % to be larger than the expected limit.
Figure 9.1 shows the post-fit distribution of the μe signal region and the composition of the μe top

quark control region. Post-fit means here, that the best fit values for all nuisance parameters and the signal strength are used for the shown distribution. Please note, that the signal is, in contrast to all pre-fit plots, stacked with the backgrounds. The plots demonstrate that the observed data are described well by the post-fit signal and background model.
Table 9.5[1] gives the post-fit event yields meaning that the best fit values for all nuisance parameters and the signal strength are used. It lists the event yields of all processes considered by the fit for the μe top quark control region, the μe signal region and the two highest bins in the BDT score of the μe signal region. It can be seen that the observed data are in good agreement with the fitted event yields of background and signal. The large uncertainties on the signal yield are due to the large uncertainties on the signal strength μ.
The so called pull-plot for the fit performed in the μe final state is given in figure 9.2. A pull-plot illustrates the impact of all considered nuisance parameters[2] θ on the best fitted signal strength $\hat{\mu}$ and the effect of the fit on the nuisance parameters themselves. The nuisance parameters are ranked with respect to their post-fit impact on $\hat{\mu}$ from large impact on the top to a small impact on the bottom. The impact of a nuisance parameter on the signal strength is measured by changing the nominal value of one nuisance parameter at a time to its nominal value plus (minus) its upper (lower) uncertainty and performing a fit with each shifted value of the nuisance parameter. Then the difference in $\hat{\mu}$ between the fit with the nominal and shifted value of the nuisance parameter, denoted by $\Delta\hat{\mu}$, is calculated. This is done once with the pre-fit and post-fit values of the nuisance parameters corresponding to the pre- and post-fit impact of the nuisance parameter. A large value of $\Delta\hat{\mu}$ corresponds to a large impact of this nuisance parameter on the fitted signal strength. The color code of the nuisance parameters in the labels of the y-axis illustrates the nuisance parameters' origin.
A pull-plot illustrates furthermore the impact of the fit on the nuisance parameters θ themselves by showing their so called pulls and constraints. A nuisance parameter is said to be pulled if the best fit value of the nuisance parameter $\hat{\theta}$ is shifted to another value than the nominal pre-fit value, denoted by θ_0. The pull-plots show the pull of each systematic and a relative scale by dividing $\hat{\theta}-\theta_0$ by the pre-fit uncertainty of the nuisance parameter which is given by $\Delta\theta$. A nuisance parameter is said to be constrained if its pre-fit uncertainty is larger than its post-fit uncertainty. The pull-plot for this fit shows that none of the considered nuisance parameters is pulled by the fit to a best-fit value lying outside of its pre-fit uncertainties. It can be seen that nevertheless one of the nuisance parameters is pulled relatively strong compared to all other nuisance parameters. This nuisance parameter describes the systematic uncertainty of the shape of the distribution of the fake-lepton background (labeled by *FakeHistoSysMuE*). The prediction of this background is known to be not precise, as its modelling is very challenging.
The fit is able to constrain the uncertainties of some nuisance parameters due to the observed data. This can be seen in the pull-plot by comparing the pre-fit uncertainties of the nuisance parameters, denoted by the red error bars, with the post-fit uncertainties given by the black error bars. It can be seen that one nuisance parameter, the normalization of the top quark background (labeled by *ATLAS_norm_Top_MuE*), gets constrained a lot. The normalization of the top quark background is a free parameter in the fit that is allowed to float in a relatively wide range, hence the large

[1] Please note that the systematic uncertainties on individual background contributions can be of the same size, or even larger than the statistical and systematical uncertainties of the total background, as the fit might not be able to constrain the contribution of an individual background as much as the total background contribution due to correlations between the different background contributions.

[2] The term *nuisance parameters* denotes here all systematic uncertainties and the free-floating normalization factors considered in the fit.

pre-fit uncertainty. The fit is able to constrain this parameter significantly due to the top quark control region that was solely introduced for this purpose. In a nutshell, the pull-plot shows no worrisome feature making the fit results trustworthy.

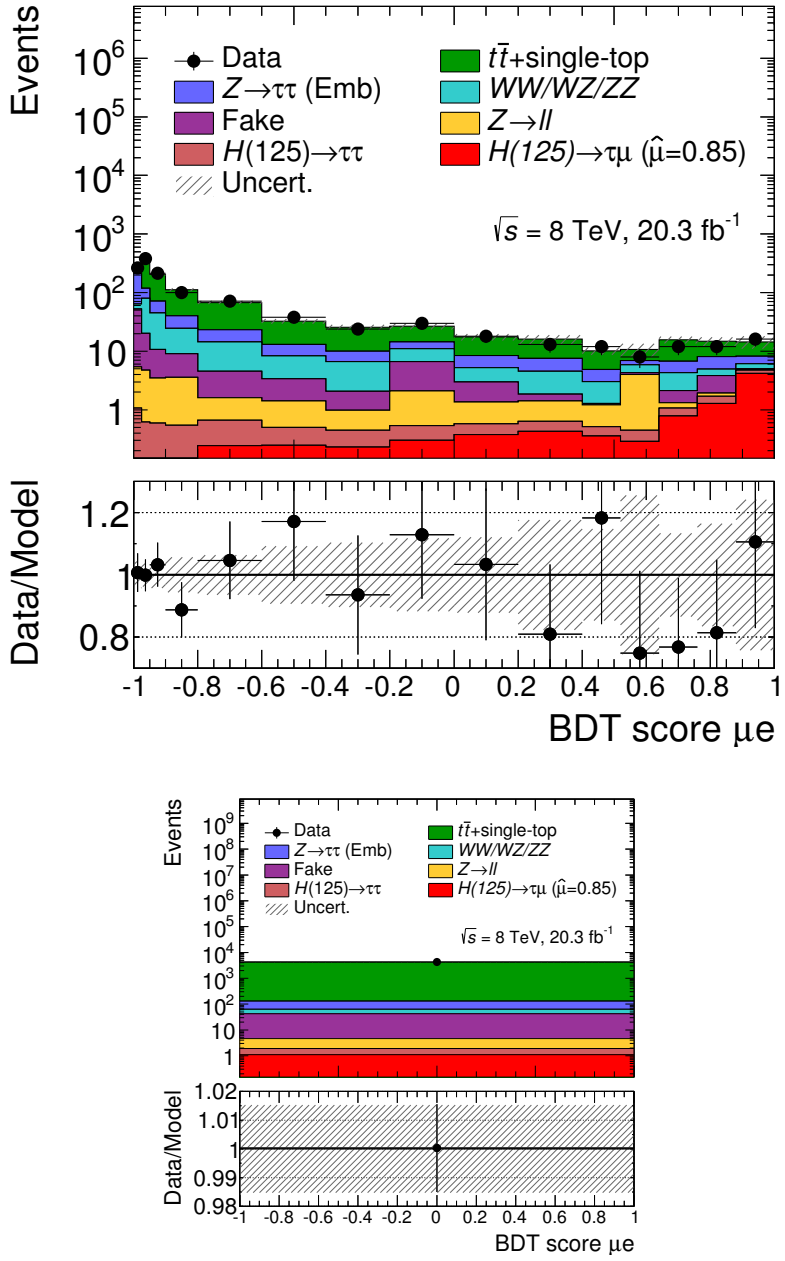


Figure 9.1: Post-fit distribution of the BDT score in the μe signal region (top) and post-fit composition of the μe top quark control region (bottom) in the μe final state. The shown uncertainties correspond to the full systematic and statistic uncertainties of signal and background. Best fit values are used for all shown distributions. Please note that these plots show signal and background stacked on each other in contrast to the pre-fit plots.

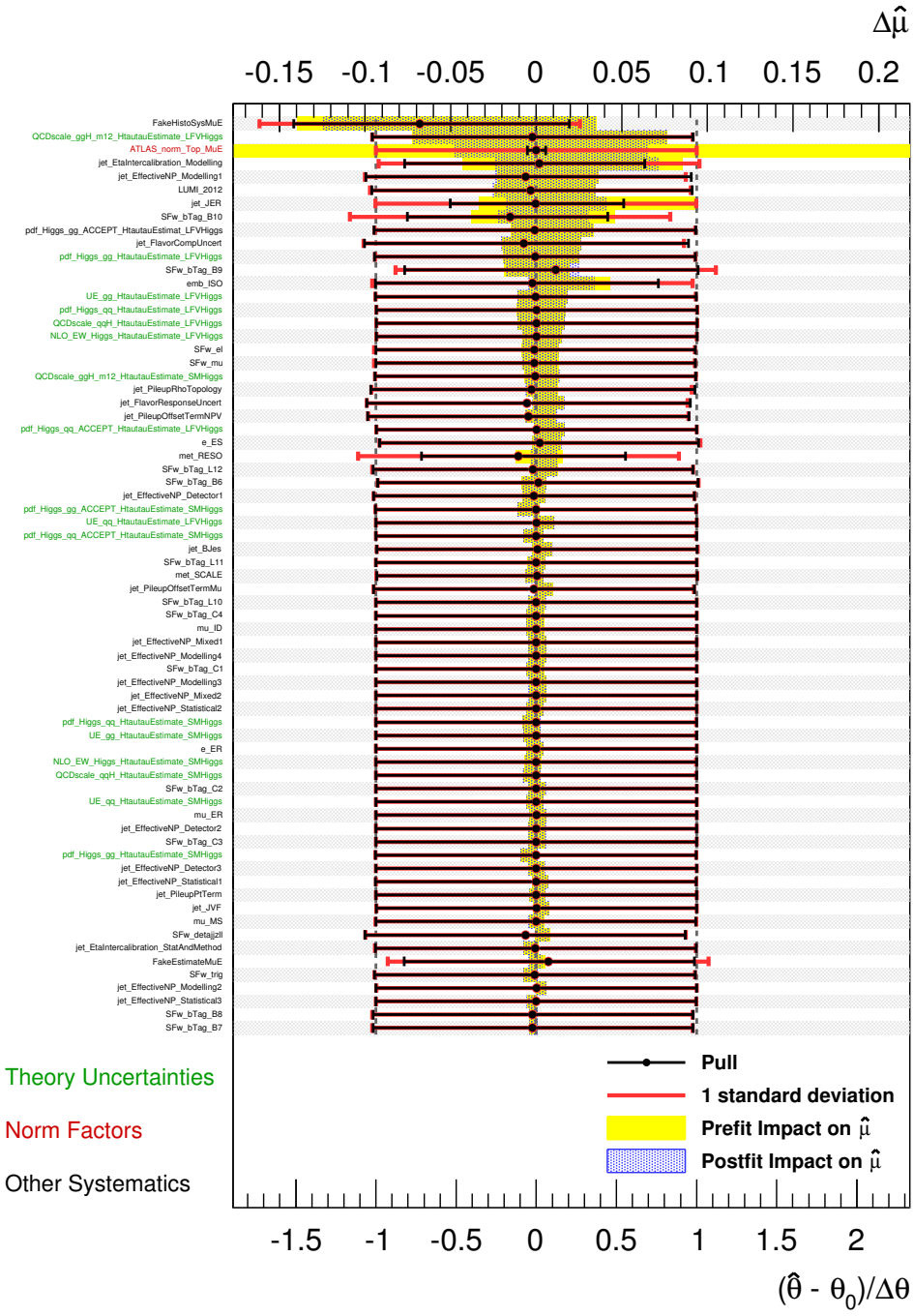


Figure 9.2: Pull-plot of the fit performed in the μe final state. This plot has two different x-axes, one on the top and one on the bottom. The x-axis given on the top quantifies the impact of each listed nuisance parameter θ on the best-fit value of the signal strength $\hat{\mu}$ given by the yellow and blue dotted bands corresponding to the pre- and post-fit impact, respectively. The lower x-axis quantifies the impact of the fit on the nuisance parameters themselves. The black dots represent the so called pull of each nuisance parameter given by the difference of the best fit value $\hat{\theta}$ and the pre-fit value θ_0 normalized to the pre-fit uncertainties $\Delta\theta$. The red and black error bars represent the pre- and post-fit uncertainties of the nuisance parameter, respectively.

9.2.2 Fit for $H \to \tau\mu$ in the $\mu\mu$ Final State

Expected Sensitivity

The expected sensitivity from fits to various Asimov datasets is summarized in table 9.6. The fit is capable to retrieve the input value of the signal strength μ well without showing any hints for a bias. It can be seen that the $\mu\mu$ final state is less sensitive that the previously discussed μe final state, since the expected significance is lower and the expected upper limit on the branching ratio of the $H \to \tau\mu$ decay, given in table 9.7, is weaker. This is expected due to the overwhelming $Z \to \mu\mu$ background in this final state yielding a smaller signal to background ratio.

μ in Asimov data	$\hat{\mu}$	expected significance
0	0.00 ± 1.14	0.01
1	1.00 ± 1.33	0.79
5	5.00 ± 1.98	2.98

Table 9.6: Performance of the fit in the $\mu\mu$ final state on Asimov data with a different input value of the signal strength μ.

$+2\,\sigma$	6.3 %
$+1\,\sigma$	4.2 %
Median	**2.7** %
$-1\,\sigma$	1.9 %
$-2\,\sigma$	1.3 %

Table 9.7: The expected upper limit and its uncertainties on the branching ratio for the decay $H \to \tau\mu$ at a confidence level of 95 % derived from the fit in the $\mu\mu$ final state.

Results in Data

No hints for the LFV decay $H \to \tau\mu$ are observed. The results of the fit performed on the observed data in the $\mu\mu$ final state are given in table 9.8. The signal strength is fitted to a negative value of $\hat{\mu} = -3.5 \pm 1.00$. This is due to a lack of observed data in the highest bin of the BDT score in the $\mu\mu$ signal region compared to the background predictions. This lack of observed data explains as well the fact that the observed limit is with 1.5 % much stricter than the expected limit of 2.7 %. The observed limit is nevertheless still covered by the 2σ uncertainty of the expected limit. Table 9.9 gives the post-fit event yields in the $Z \to \mu\mu$ control region, the $\mu\mu$ signal region and the

$\hat{\mu}$	observed significance	observed limit BR($H \to \tau\mu$)
-3.5 ± 1.00	$-2.6\,\sigma$	1.5 %

Table 9.8: Best fit value for μ, observed significance and upper limit on BR($H \to \tau\mu$) at 95 % confidence level for the fit in the μe final state.

two highest bins of the BDT score in this signal region. The negative event yields for the $H \to \tau\mu$ signal correspond to the negative best fit value for the signal strength caused by the deficit of data compared to the background estimate in the highest bin of the BDT score. In this table, it can be

observed how the fit compensates for the over-estimation of the background in the highest bin of the BDT score: The too large post-fit background estimation of 8.6 events is corrected by a fitted negative signal yield of -7 events, such that the combined event yield of signal and background is in good agreement with the observed data of 1. Another effect of this downward fluctuation is that the observed limit is much stricter than the expected limit.
The pull-plot for this fit is shown in figure 9.4. Please note that not all post-fit uncertainties of the nuisance parameters are shown in this plot, as their numerical evaluation failed. This affects obviously the post-fit impact on $\hat{\mu}$ and therefore the ranking of the nuisance parameters, since the post-fit uncertainties are used for the derivation of the post-fit impact on $\hat{\mu}$. The ranking and post-fit impact on $\hat{\mu}$ is therefore not trustworthy for all nuisance parameters that are affected by this (nuisance parameters with no or only an upper or lower black error bar). These nuisance parameters were shown nevertheless since their pre-fit impact on $\hat{\mu}$ and the pull is calculated correctly. Unfortunately, it was not possible to solve this technical issue within the timescale of this masters thesis. The shown properties of nuisance parameters that are not affected by this should be trustworthy.
The pull-plot allows nevertheless to draw some conclusions: First of all, it shows that the nuisance parameters have a much bigger pre- and post-fit influence on the best fit value of μ, compared to the μe final state. Additionally, it can be seen that multiple nuisance parameters are pulled by the fit from their nominal value, but none of the nuisance parameters is pulled to a best fit value that lies significantly outside of its pre-fit $1\,\sigma$ uncertainty. The presence of the pulls is an artifact of the fit's effort to absorb the observed discrepancies between the background estimates and data in the two poorly described bins that were discussed in section 7.5.2. This discrepancies are either some very unfortunate fluctuations or caused by a minor mismodeling. Unfortunately, it was beyond the time scale of this master thesis to investigate the mechanisms possibly causing these discrepancies and to cure the analysis from them.

Process/Category	SR			CR
	Total	2nd last bin	Last bin	Total
$Z \to \mu\mu$	4280 ± 60	15.2 ± 1.8	2.3 ± 0.5	85110 ± 300
Top quark	203 ± 24	3.3 ± 0.5	1.44 ± 0.34	460 ± 60
$Z \to \tau\tau$ (Emb)	86 ± 4	4.7 ± 0.4	1.91 ± 0.27	139 ± 6
WW/WZ/ZZ	97 ± 4	1.53 ± 0.14	0.249 ± 0.035	603 ± 15
Fake-leptons	90 ± 18	6.6 ± 2.7	2.3 ± 0.9	155 ± 31
VBF $H \to \tau\tau$	3.11 ± 0.18	0.284 ± 0.028	0.26 ± 0.04	1.17 ± 0.06
ggH $H \to \tau\tau$	2.1 ± 0.6	0.17 ± 0.05	0.15 ± 0.05	1.6 ± 0.4
VBF $H \to \tau\mu$	-23 ± 5	-3.8 ± 0.9	-4.6 ± 1.3	-6.8 ± 1.6
ggH $H \to \tau\mu$	-17 ± 5	-3.0 ± 1.0	-2.4 ± 0.8	-13 ± 4
Total Bkd.	4760 ± 60	31.7 ± 3.4	8.6 ± 1.6	86470 ± 290
Data	4695	28	1	86487

Table 9.9: Expected event yield and data in the $\mu\mu$ signal region and in the $Z \to \mu\mu$ control region (CR) after the fit in the $\mu\mu$ final state. The composition of the last and 2nd last bin of the BDT score (two highest bins in the BDT score) of the $\mu\mu$ signal region is given additionally. The uncertainties of the individual processes are only systematic uncertainties, while the given uncertainty on the total background (Bkd.) correspond to systematic and statistical uncertainties.

The previously discussed pulls of the nuisance parameters are nevertheless a reasonable behavior of the fit allowing it to bring the background and signal estimates into agreement with the observed data. This can be seen in table 9.9 that shows a good agreement between the sum of the background estimates and the negative signal yield and the observed data in all listed bins and regions. This is additionally illustrated in figure 9.3 showing the post-fit distribution of the BDT score in the $\mu\mu$ signal region and the composition of the $Z \to \mu\mu$ control region. Please note that these plots do not show the fitted signal due to its negative signal yield. This causes that the post-fit distribution of the BDT score in the $\mu\mu$ signal region appears to have a tension between observed data and background and signal estimates in the highest bin of the BDT score. Considering the negative fitted event yield, as previously discussed, reveals that observed data and post-fit background and signal estimates are in this bin in good agreement.
Concluding, the pulls of the nuisance parameters should be studied and understood to trust the fit results completely, but the behavior of the fit is nevertheless reasonable, since it absorbs the tensions between the background and signal modeling and the observed data. This and the fact that there are no pulls of nuisance parameters that are significantly larger than their $1\,\sigma$ pre-fit uncertainties allows to have some confidence in the fit results.

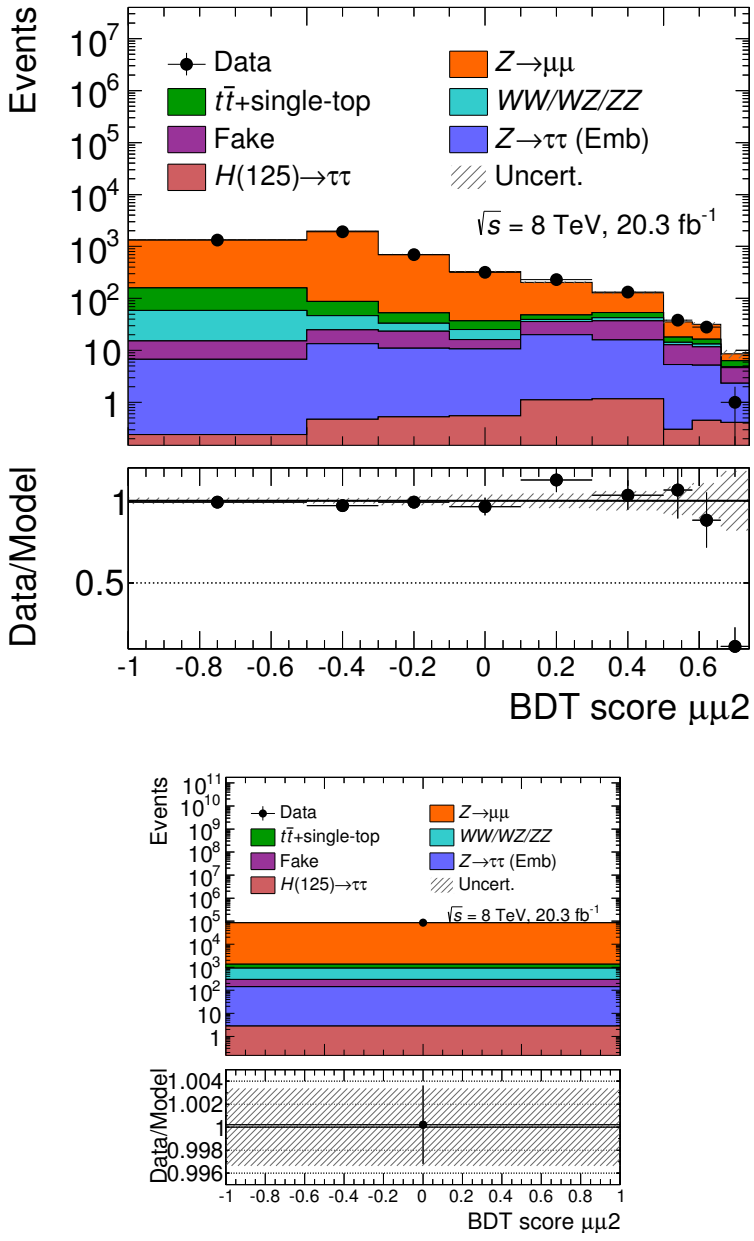


Figure 9.3: Post-fit BDT distributions of $\mu\mu$ signal region and the $Z \to \mu\mu$ control region for the fit in the $\mu\mu$ final state. The fitted signal is not shown due to the negative fitted $\hat{\mu}$. The given uncertainties correspond to the systematic and statistic uncertainties of the background.

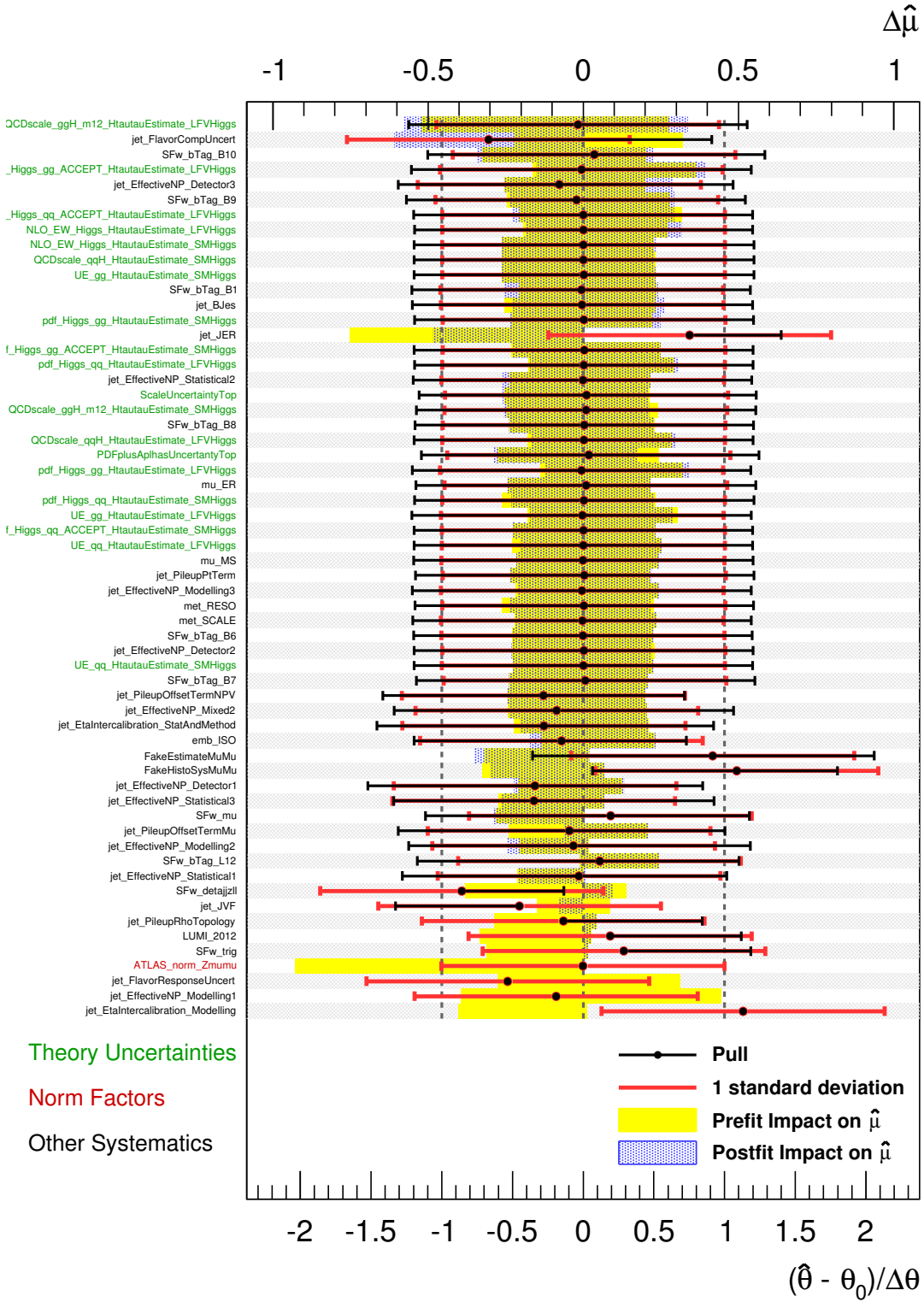


Figure 9.4: Pull-plot of the $\mu\mu$ final state. A detailed discussion of this pull-plot is given in the main text. Pull-plots in general are explained in the caption of figure 9.2.

9.2.3 Fit for $H \to \tau e$ in the $e\mu$ Final State

Expected Sensitivity

The fit in the $e\mu$ final state for the $H \to \tau e$ signal is tested using Asimov data with different input values of the signal strength μ. Table 9.10 gives the performance of the fit on the tested Asimov datasets. It shows that the fit is capable to recover the input value of μ well without showing a bias. The expected upper limit on the $H \to \tau e$ signal of 1.5 % and its uncertainties are given in table 9.11.

μ in Asimov data	$\hat{\mu}$	expected significance
0	0.00 ± 0.73	-0.01
1	1.00 ± 0.85	1.08
5	4.99 ± 1.32	4.20

Table 9.10: Performance of the fit in the $e\mu$ final state on Asimov data with a different input value of the signal strength μ. The best fit value $\hat{\mu}$ of the signal strength and the expected significance is stated for each input value of μ.

$+2\,\sigma$	3.1 %
$+1\,\sigma$	2.2 %
Median	**1.5** %
$-1\,\sigma$	1.0 %
$-2\,\sigma$	0.8 %

Table 9.11: The upper limit and its uncertainties on the branching ratio (BR) for the decay $H \to \tau e$ at a confidence level of 95 % derived from the $e\mu$ final state.

Results in Data

No hints for the LFV decay $H \to \tau e$ are observed. The information on the $H \to \tau e$ signal extracted from the fit in the $e\mu$ final state is summarized in table 9.12. The best-fit signal strength

best fit for μ	observed significance	observed limit BR($H \to \tau\mu$)
0.1 ± 1.0	$0.1\,\sigma$	1.9 %

Table 9.12: Best fit value for μ, observed significance and upper limit on BR($H \to \tau e$) at 95 % confidence level for the fit in the $e\mu$ final state.

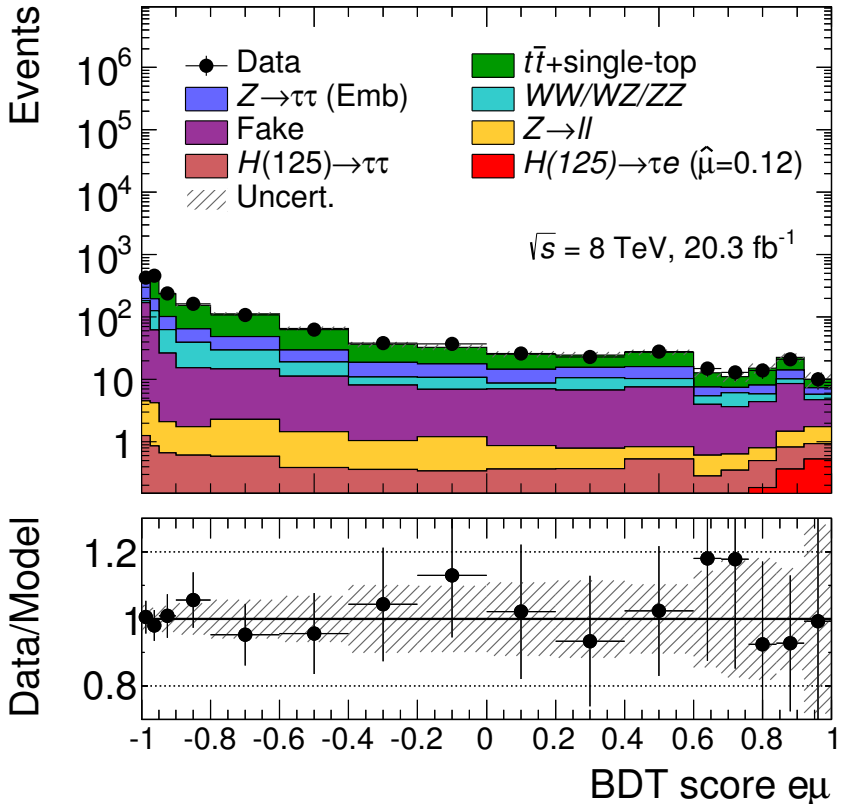
is $\hat{\mu} = 0.1 \pm 1.0$ and is therefore in good agreement with the SM prediction of $\mu = 0$, which is highlighted as well by the small observed significance of $0.1\,\sigma$. The observed upper limit on the branching ratio of the decay $H \to \tau e$ is with 1.9 % in good agreement with the expected upper limit of 1.5 %, as it is well within the $1\,\sigma$ uncertainty of the expected limit.
Figure 9.5 shows the post-fit distribution of the BDT score in the $e\mu$ signal region and the composition of the $e\mu$ top quark control region using the best fit values for all parameters. These plots pronounce a good agreement between observed data and the fitted signal and background estimates.

The pull-plot for this fit is given in 9.6 showing no large pulls of the nuisance parameters. The only nuisance parameter that is constrained significantly is the normalization of the top quark background, which is expected due to its wide pre-fit parameter range. The two uncertainties on the fake-lepton background estimate have the largest impact on the fitted signal strength. This

Process/Region	$e\mu$ Signal Region			$e\mu$ Top CR
	Total	2nd last bin	Last bin	Total
Top quark	740 ± 50	8.5 ± 1.6	2.8 ± 0.6	4730 ± 70
$Z \to \tau\tau$ (Emb)	397 ± 22	3.9 ± 0.5	1.56 ± 0.20	96 ± 7
WW/WZ/ZZ	183 ± 17	1.7 ± 1.3	1.0 ± 0.4	23.3 ± 2.8
Fake-leptons	340 ± 50	7.1 ± 1.8	2.9 ± 3.2	117 ± 18
$Z \to ll$	17.1 ± 2.2	0.64 ± 0.21	0.82 ± 0.24	2.9 ± 0.8
VBF $H \to \tau\tau$	3.90 ± 0.28	0.30 ± 0.04	0.31 ± 0.05	0.48 ± 0.05
ggH $H \to \tau\tau$	3.6 ± 1.0	0.15 ± 0.05	< 0.1	0.50 ± 0.15
VBF $H \to \tau e$	1 ± 8	0.2 ± 1.7	0.4 ± 3.5	0.1 ± 1.0
ggH $H \to \tau e$	1 ± 8	0.2 ± 1.5	0.1 ± 1.1	0.2 ± 1.3
Total Bkd.	1680 ± 40	22.3 ± 3.4	9.5 ± 2.8	4970 ± 70
Data	1684	21	10	4971

Table 9.13: Expected event yield and data in the $e\mu$ signal region and in the $e\mu$ top quark control region (CR) after the fit in the $e\mu$ final state. The composition of the last and 2nd last bin of the BDT score (two highest bins in the BDT score) of the $e\mu$ signal region is given additionally. The uncertainties of the individual processes are only systematic uncertainties, while the given uncertainty on the total background (Bkd.) corresponds to systematic and statistical uncertainties.

can be understood by the post-fit composition of the backgrounds in the highest bins of the BDT score of the $e\mu$ signal region, given together with the event yield of the $e\mu$ signal and top quark control region in table 9.13. These two bins contain most information on the signal. The fake-lepton background is one of the dominant backgrounds in these bins, contributing roughly one third of the total background. The systematic uncertainty on the fake-lepton background is in these bins relatively large. This is especially true for the highest bin in the BDT score meaning that a pull of its systematic can either fake or mask a signal, hence the large impact of its uncertainties on $\hat{\mu}$. These systematic uncertainties are not pulled significantly meaning that it is unlikely that they faked or masked a signal in this fit, but their large impact can limit the sensitivity of the analysis. This issue could potentially be solved by improving the estimate of the fake-background with respect to its systematic uncertainties or by reducing the contribution of the fake-lepton background significantly. Both would yield extended studies that could not be performed in the timescale of this thesis. The pull-plot of this final states does not show any worrisome behavior of the fit besides the previously discussed possible limitation of the sensitivity due to the large impact of the fake-lepton background on the fitted signal strength.
The post-fit event yields, shown in table 9.13, are in good agreement with the post-fit background and signal estimates. The large uncertainties on the post-fit signal yield can be explained by the large uncertainty on $\hat{\mu}$.



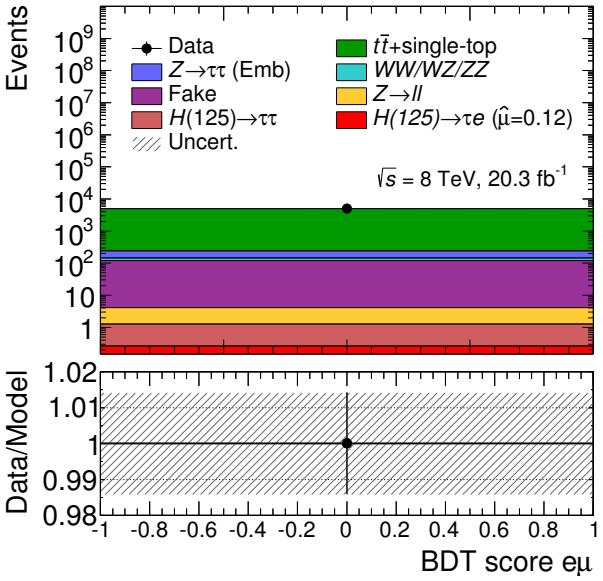


Figure 9.5: Post-fit distribution of the BDT score in the $e\mu$ signal region (top) and post-fit composition of the $e\mu$ top quark control region (bottom) after the fit in the $e\mu$ final state. The uncertainty bands correspond to the statistic and systematic uncertainties of signal and background.

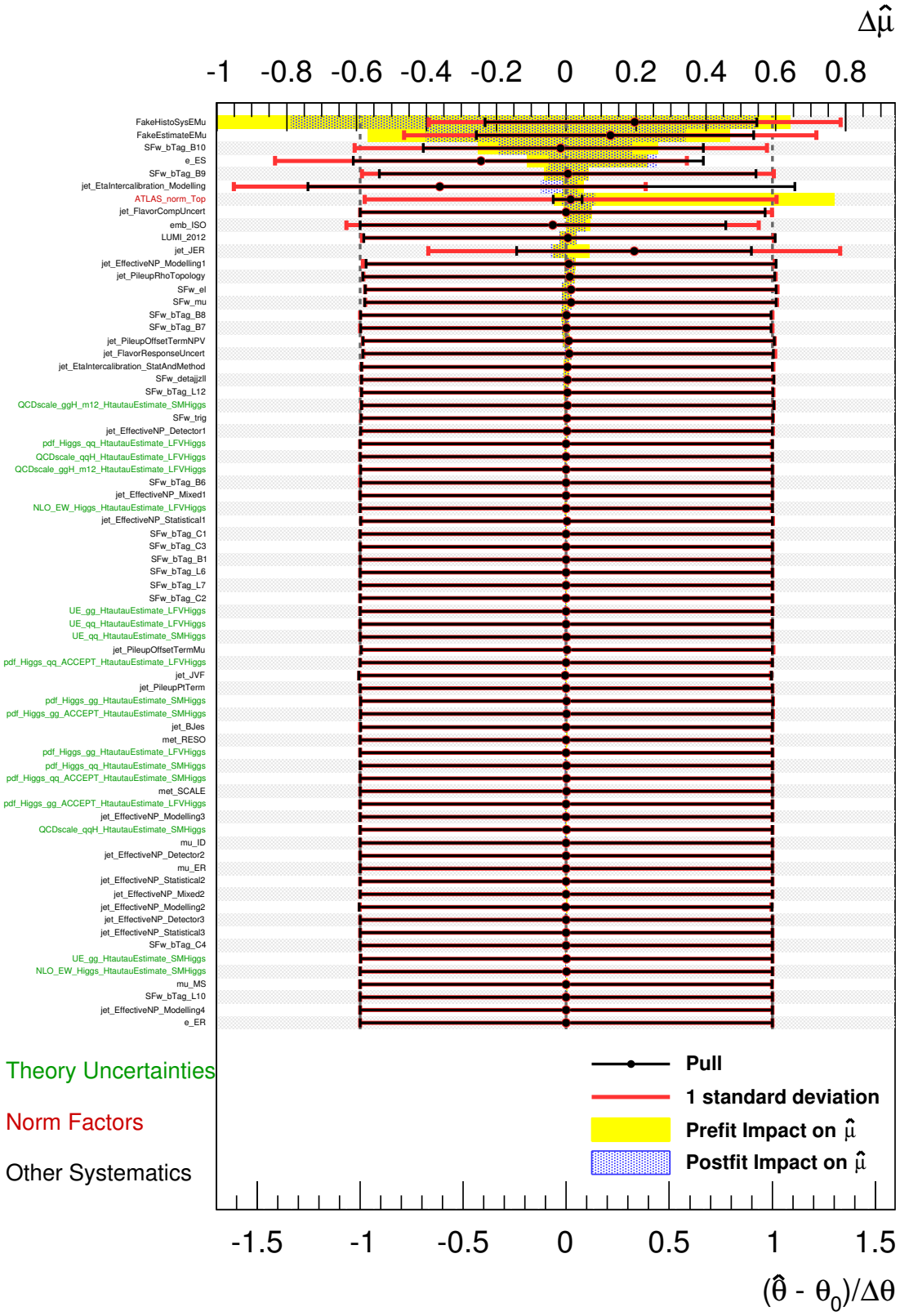


Figure 9.6: Pull-plot of the fit for the $H \to \tau e$ signal in the $e\mu$ final state. A detailed discussion of this pull-plot is given in the main text. Pull-plots in general are explained in the caption of figure 9.2.

9.2.4 Fit for $H \to \tau e$ in the ee Final State

Expected Sensitivity

The behavior of the fit for the decay $H \to \tau e$ in the ee final state on Asimov data is given in table 9.14. The fit shows no bias as it is able to recover the input value of the signal strength well. The large uncertainties on $\hat{\mu}$ and the small expected significance indicate both the small sensitivity of this final state compared to the other final states. This is reasonable since both same flavor final states are less sensitive than the different flavor final state due to their large $Z \to ll$ background contribution yielding a small signal to background ratio. Furthermore, it was already seen in the selection of the binning for the BDT score of the same flavor final states, described in section 7.5.2, that the binning of the ee final state yields a smaller Asimov significance for the highest bin of the BDT score than in the $\mu\mu$ final state. The small sensitivity of this final state is reflected as well by the expected upper limit on the $H \to \tau e$ decay of 3.6 % that is given with its uncertainties in table 9.15.

μ in Asimov data	$\hat{\mu}$	expected significance
0	0.00 ± 1.47	0.00
1	1.00 ± 1.675	0.62
5	5.00 ± 2.42	2.42

Table 9.14: Performance of the fit in the ee final state on Asimov data with a different input value of the signal strength μ.

$+2\,\sigma$	8.6 %
$+1\,\sigma$	5.6 %
Median	**3.6** %
$-1\,\sigma$	2.4 %
$-2\,\sigma$	1.7 %

Table 9.15: The upper limit and its uncertainties on the branching ratio (BR) of the decay $H \to \tau e$ at a confidence level of 95 % derived from the ee final state.

Results in Data

No hints for the LFV decay $H \to \tau e$ are observed in this final state. Table 9.16 gives the results of the fit for the $H \to \tau e$ signal in the ee final state. The fitted signal strength of -0.5 ± 1.7 is

best fit for μ	observed significance	observed limit BR($H \to \tau e$)
-0.5 ± 1.7	$-0.3\,\sigma$	3.3 %

Table 9.16: Best fit value for μ, observed significance and upper limit on BR($H \to \tau e$) at 95 % confidence level for the fit in the ee final state.

negative but is within its uncertainties well compatible with the SM prediction of $\mu = 0$, which is confirmed by the observed significance of $-0.3\,\sigma$. The observed upper limit on the branching ratio of the decay $H \to \tau e$ of 3.3 % is a little better than the expected limit of 3.6 % which is reasonable

due to the negative value of $\hat{\mu}$. It is nevertheless still covered by the 1σ uncertainties of the expected limit.
A good agreement of observed data and post-fit signal and background estimates can be seen in the post-fit plots given in figure 9.7. They show the post-fit distribution of the ee signal region and the composition of the $Z \to ee$ control region after the fit in the ee final state. These plots do not show any signal due to its negative expected event yield. The fitted negative event yield of the signal reduces the predicted event yield in the highest bins of the BDT score slightly making the agreement of the data and predicted event yield in the two highest bins of the BDT score of the ee signal region even better than shown in the plots.

Process/Category	ee Signal Region			$Z \to ee$ CR
	Total	2nd last bin	Last bin	Total
$Z \to ee$	3050 ± 60	23.1 ± 2.5	1.79 ± 0.27	60530 ± 270
Top quark	125 ± 16	5.4 ± 0.9	1.8 ± 0.4	350 ± 50
$Z \to \tau\tau$ (Emb)	27.7 ± 2.8	2.13 ± 0.27	1.09 ± 0.18	84 ± 6
WW/WZ/ZZ	57.6 ± 3.0	0.68 ± 0.09	0.50 ± 0.08	433 ± 11
Fake-leptons	51 ± 26	3.8 ± 1.6	0.9 ± 1.8	150 ± 80
VBF $H \to \tau\tau$	1.65 ± 0.10	0.186 ± 0.019	0.27 ± 0.04	0.86 ± 0.05
ggH $H \to \tau\tau$	0.86 ± 0.23	< 0.1	0.14 ± 0.04	1.09 ± 0.29
VBF $H \to \tau e$	-2 ± 6	-0.3 ± 1.0	-0.5 ± 1.6	-0.6 ± 1.9
ggH $H \to \tau e$	-1 ± 4	-0.2 ± 0.7	-0.3 ± 1.1	-1 ± 4
Total Bkd.	3310 ± 60	35.4 ± 3.5	6.5 ± 2.3	61550 ± 260
Data	3309	29	6	61544

Table 9.17: Expected event yield and data in the ee signal region and in the $Z \to ee$ top quark control region (CR) after the fit in the ee final state. The composition of the last and 2nd last bin of the BDT score (two highest bins in the BDT score) of the ee signal region is given additionally. The uncertainties of the individual processes are only systematic uncertainties, while the given uncertainty on the total background (Bkd.) correspond to systematic and statistical uncertainties.

The pull-plot for this fit is given in figure 9.8. It shows that none of the considered nuisance parameters is pulled significantly by the fit. The nuisance parameter with the biggest influence on $\hat{\mu}$ is the systematic for the shape of the distribution of the fake-lepton background. This can be understood by the large systematic uncertainty of the fake-lepton background in the highest bin of the BDT score of the ee signal region given in table 9.17 that can fake or mask a signal in the highest and most sensitive bin, hence the large impact of the systematics of the fake-lepton estimate on $\hat{\mu}$.
Table 9.17 gives additionally the post-fit event yields in the $Z \to ee$ control region, the ee signal region and the two highest bins in the BDT score of the ee signal region. The negative event yields for the $H \to \tau e$ signal and the the negative best fit value for the signal strength are caused by the deficit of data compared to the background estimate in the highest bins of the BDT score. All predicted event yields agree well with the observed data. The fit is therefore able to compensate for shortage of predicted events prior to the fit in the ee signal region. By comparing the pre-fit event yields that were given in table 7.12 it can be seen that this is achieved by increasing the event yield of the dominant $Z \to ee$ background. The fit is able to do this by pulling systematic uncertainties with a large impact on the expected event yield slightly. It cannot pull the $Z \to ee$ normalization factor since this is constrained strongly by the $Z \to ee$ control region. Table 9.18

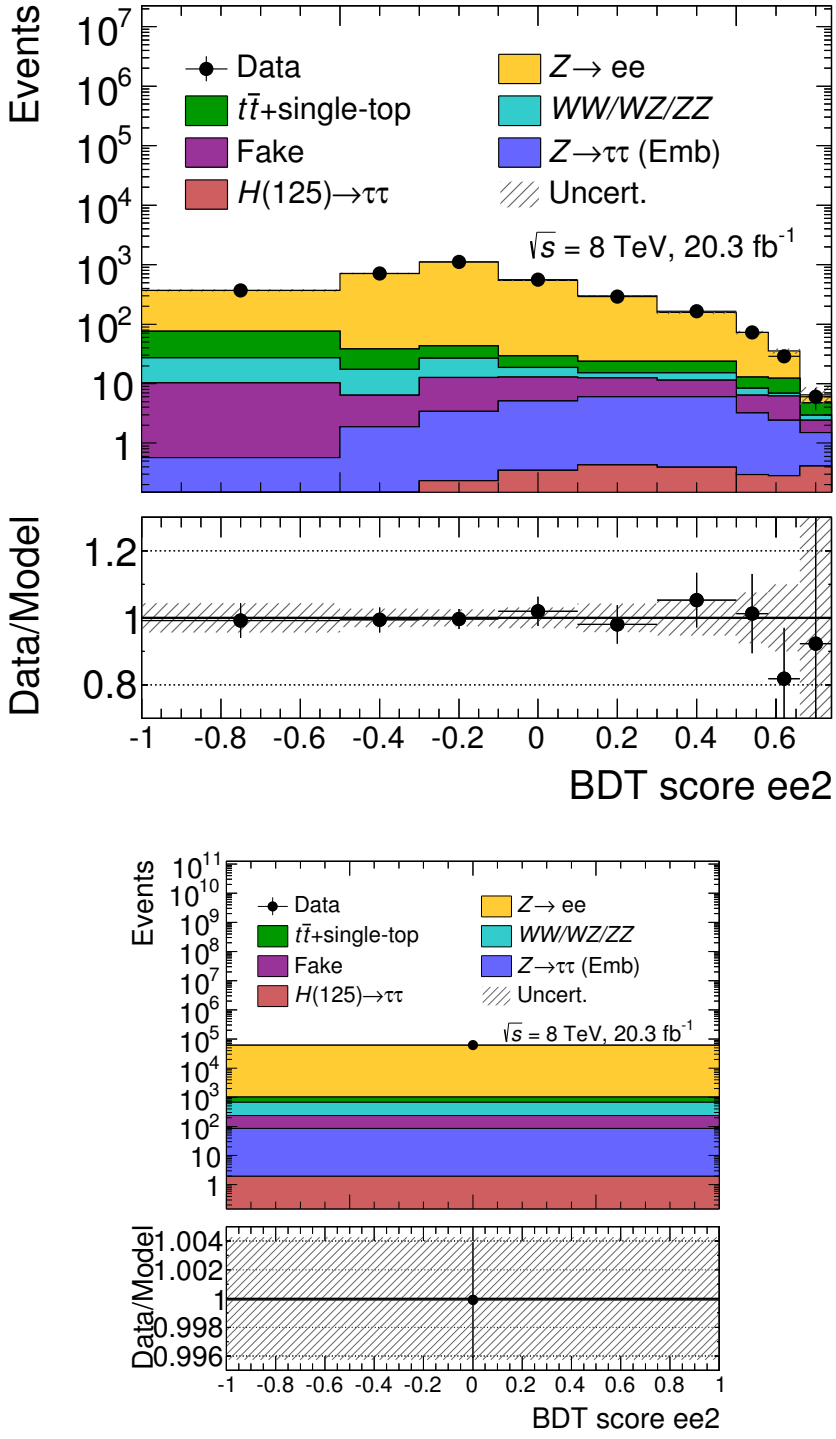


Figure 9.7: Distribution of the BDT score in the *ee* signal region (left) and composition of the $Z \to ee$ control region (right) using the best-fit parameters of the *ee* final state. The fitted signal is not shown due to the negative fitted signal strength. The shown uncertainties correspond to the systematic and statistic uncertainties of the background.

gives the three systematic uncertainties that have the biggest influence on the event yield of the $Z \to ee$ background estimate in the ee signal region. Their impact on the ee signal region is enough to compensate easily the shortage of the $Z \to ee$ pre-fit estimate. The pull-plot given in figure 9.8 shows that all these systematics are pulled sightly towards larger values than their nominal value, while the nuisance parameter describing the normalization of the $Z \to ee$ background is not pulled at all.

Systematic uncertainty	Change of event yield by a shift of $-1\,\sigma$	$+1\,\sigma$
e_ES	$-3\,\%$	$+24\,\%$
jet_EtaIntercalibration_Modelling	$-13\,\%$	$+16\,\%$
jet_JER	$-14\,\%$	$+15\,\%$

Table 9.18: The three systematics with the biggest influence on the event yield of the $Z \to ll$ background in the signal region of the ee final state and their effect on the event yield by pulling them to the edges of their 1σ confidence interval.

Figure 9.8: Pull-plot of the fit for the $H \to \tau e$ in the ee final state. A detailed discussion of this pull-plot is given in the main text. Pull-plots in general are explained in the caption of figure 9.2.

9.3 Combined Fits

This section presents combined fits for the $H \to \tau\mu$ and $H \to \tau e$ signal. A simultaneous fit for the $H \to \tau\mu$ ($H \to \tau e$) signal is performed in the μe and $\mu\mu$ ($e\mu$ and ee) final state. All systematic uncertainties are assumed to be fully correlated in all final states considered in each simultaneous fit, except the systematic uncertainties of the fake-lepton estimate, which are assumed to be uncorrelated. The normalization of the $Z \to ee$, $Z \to \mu\mu$ and top quark background is only allowed to float in those final states that have a dedicated control region for that background.
The main idea behind these simultaneous fits is to increase the sensitivity for the signal by evaluating all available information simultaneously. This limits furthermore the influence of statistical fluctuations in data and in the background modeling on the nuisance parameters and on the fitted signal strength due to the larger and more constraining datasets.

9.3.1 Simultaneous Fit for $H \to \tau\mu$ in the $\mu\mu$ and μe Final State

This fit is carried out in the μe signal region, the μe top quark control region, the $\mu\mu$ signal region and the $Z \to \mu\mu$ control region.

Expected Sensitivity

The performance of this simultaneous fit is tested on a set Asimov data with different input values for the signal strength μ. Table 9.19 summarizes the results of these tests. It shows that the fit is able to recover the input value of μ well without a bias. The expected significance of this

μ in Asimov data	fitted μ	expected significance
0	0.00 ± 0.61	-0.01
1	1.00 ± 0.735	1.52
5	5.00 ± 1.21	6.1

Table 9.19: Performance of a combined fit for $H \to \tau\mu$ on Asimov data.

simultaneous fit is larger than the expected significance of the previously shown fits on the individual final states. The expected upper limit on the branching ratio of the $H \to \tau\mu$ signal, given with its uncertainty in table 9.20, of 1.4 % is stricter than the ones derived from the individual fit in the μe and $\mu\mu$ final state of 1.7 % and 2.7 %, respectively.

$+2\,\sigma$	3.0 %
$+1\,\sigma$	2.1 %
Median	**1.4** %
$-1\,\sigma$	1.0 %
$-2\,\sigma$	0.7 %

Table 9.20: The upper limit and its uncertainties on the branching ratio for the decay $H \to \tau\mu$ at a confidence level of 95 % .

Results in data

No hints for the LFV decay $H \to \tau\mu$ are observed in the combined fit in the μe and $\mu\mu$ final state. This can be seen in table 9.21 that summarizes the fit results on the $H \to \tau\mu$ signal from the

$\hat{\mu}$	observed significance	observed limit BR($H \to \tau\mu$)
0.2 ± 0.6	$0.4\,\sigma$	1.5 %

Table 9.21: Best fit value for μ, observed significance and upper limit on the branching ratio (BR) of the decay $H \to \tau\mu$ at 95 % confidence level.

simultaneous fit. The best fit value $\hat{\mu}$ of the signal strength is fitted to 0.2 ± 0.6, with an observed significance of $0.4\,\sigma$. The fitted signal strength agrees therefore well with the SM prediction of $\mu = 0$. The observed upper limit on the branching ratio of $H \to \tau\mu$ is with 1.5 % marginally weaker than the expected limit, but still within its 1σ uncertainties.

Process/Category	μe Signal Region			μe Top Quark CR
	Total	2nd last bin	Last bin	Total
Top quark	666 ± 35	7.7 ± 1.6	6.3 ± 0.8	4150 ± 70
$Z \to \tau\tau$ (Emb)	267 ± 11	3.2 ± 0.4	2.12 ± 0.25	68 ± 4
WW/WZ/ZZ	157 ± 9	1.18 ± 0.18	1.27 ± 0.23	19.6 ± 2.9
Fake-leptons	91 ± 17	1.9 ± 0.8	0.13 ± 0.31	38 ± 7
$Z \to ll$	23.4 ± 2.0	0.21 ± 0.06	0.151 ± 0.026	2.60 ± 0.34
VBF $H \to \tau\tau$	3.07 ± 0.20	0.259 ± 0.034	0.41 ± 0.06	0.38 ± 0.04
ggH $H \to \tau\tau$	2.9 ± 0.8	0.18 ± 0.05	0.21 ± 0.06	0.45 ± 0.13
VBF $H \to \tau\mu$	1.3 ± 3.3	0.2 ± 0.5	0.8 ± 2.1	0.2 ± 0.4
ggH $H \to \tau\mu$	1.2 ± 3.1	0.2 ± 0.5	0.3 ± 0.9	0.2 ± 0.4
Total Bkd.	1210 ± 32	14.7 ± 2.2	10.6 ± 2.9	4280 ± 70
Data	1212	12	16	4283

Table 9.22: Expected event yield and data in the μe signal region and in the μe top quark control region (CR) after the simultaneous fit in the μe and $\mu\mu$ final state. The composition of the last and 2nd last bin of the BDT score (two highest bins in the BDT score) of the μe signal region is given additionally. The uncertainties of the individual processes are only systematic uncertainties, while the given uncertainty on the total background (Bkd.) corresponds to systematic and statistical uncertainties.

Table 9.22 and 9.23 give the post-fit event yields and observed data for all regions considered by the fit. The listed accumulated event yields of fitted signal and background are in good agreement with the observed data, except the highest bins of the BDT score in both signal regions. The simultaneous fit has to compromise in these two bins between the excess of data in μe signal region and the lack of data in the $\mu\mu$ signal regions and cannot match the observed data by a fitted signal that is positive or negative, respectively. This yields a tension between the observed data and the expected accumulated event yield of the fitted signal and the background estimate in the highest bin of the BDT score of the $\mu\mu$ signal region. This is as well visible in the post-fit plots of the μe and $\mu\mu$ signal region that are given in figure 9.10. Nevertheless, the post-fit background estimates and the fitted signal model the observed data well in all other bins.
The pull-plot of the combined fit is given by figure 9.11. It shows that the fit pulls some of the systematic uncertainties away from their nominal value. Only one of the nuisance parameters is pulled to a value that is not covered by its pre-fit uncertainties. This nuisance parameter is the systematic uncertainty on the shape of the fake-lepton background. The pulls are most likely caused by the $\mu\mu$ final state, since most of the strongly pulled uncertainties were already pulled by the fit in

Process/Category	$\mu\mu$ Signal Region			$Z \to \mu\mu$ CR
	Total	2nd last bin	Last bin	Total
$Z \to \mu\mu$	4240 ± 70	12.2 ± 2.1	1.5 ± 0.6	85160 ± 320
Top quark	192 ± 17	3.1 ± 0.4	0.98 ± 0.33	430 ± 40
$Z \to \tau\tau$ (Emb)	85 ± 4	4.6 ± 0.4	1.70 ± 0.24	139 ± 6
WW/WZ/ZZ	92 ± 8	1.42 ± 0.17	0.213 ± 0.035	570 ± 40
Fake-leptons	86 ± 19	5.7 ± 2.8	2.1 ± 0.9	148 ± 34
VBF $H \to \tau\tau$	2.99 ± 0.23	0.267 ± 0.034	0.22 ± 0.04	1.16 ± 0.07
ggH $H \to \tau\tau$	2.0 ± 0.6	0.16 ± 0.05	0.11 ± 0.04	1.6 ± 0.4
VBF $H \to \tau\mu$	2 ± 4	0.3 ± 0.7	0.3 ± 0.8	0.5 ± 1.3
ggH $H \to \tau\mu$	1.2 ± 3.0	0.2 ± 0.5	0.2 ± 0.4	1.0 ± 2.5
Total Bkd.	4700 ± 70	27 ± 4	6.8 ± 1.7	86460 ± 320
Data	4695	28	1	86487

Table 9.23: Expected event yield and data in the $\mu\mu$ signal region and in the $Z \to \mu\mu$ control region (CR) after the simultaneous fit in the μe and $\mu\mu$ final state. The composition of the last and 2nd last bin of the BDT score (two highest bins in the BDT score) of the $\mu\mu$ signal region is given additionally. The uncertainties of the individual processes are only systematic uncertainties, while the given uncertainty on the total background (Bkd.) correspond to systematic and statistical uncertainties.

the $\mu\mu$ final state and did not receive a significant pull in the fit of the μe final state. The modeling of the BDT score in the $\mu\mu$ final states has, as previously discussed, some issues that could not be fixed within the timescale of this thesis. The pulls are nevertheless within a reasonable range, since the best fit values of all nuisance parameters lie either within or barely outside their pre-fit uncertainties. The post-fit plots for the considered signal and control regions, given in figure 9.10 and 9.9, respectively, show additionally a good agreement between the data and the accumulated fitted signal and background estimates.

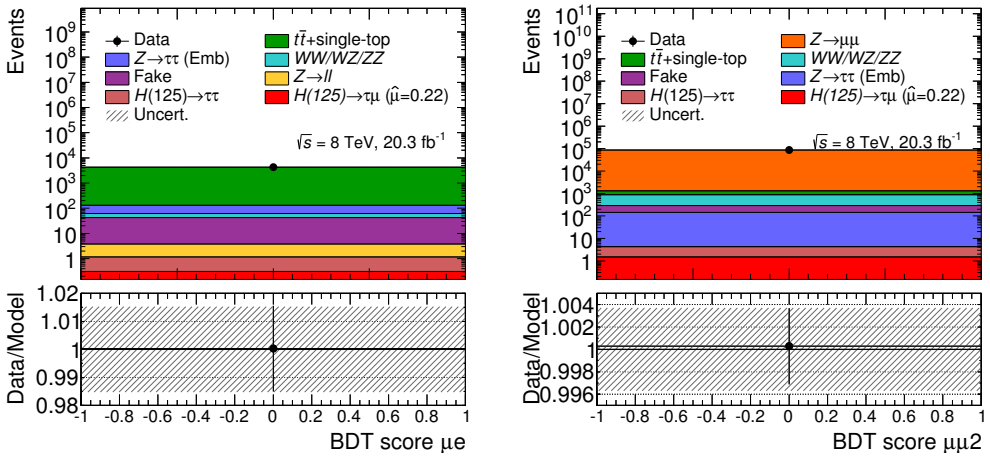


Figure 9.9: Post-fit composition of the μe top quark control region (left) and the $Z \to \mu\mu$ control region (right) for the simultaneous fit in the $\mu\mu$ and μe final state. The shown uncertainties correspond to the statistic and systematic uncertainties of signal and background.

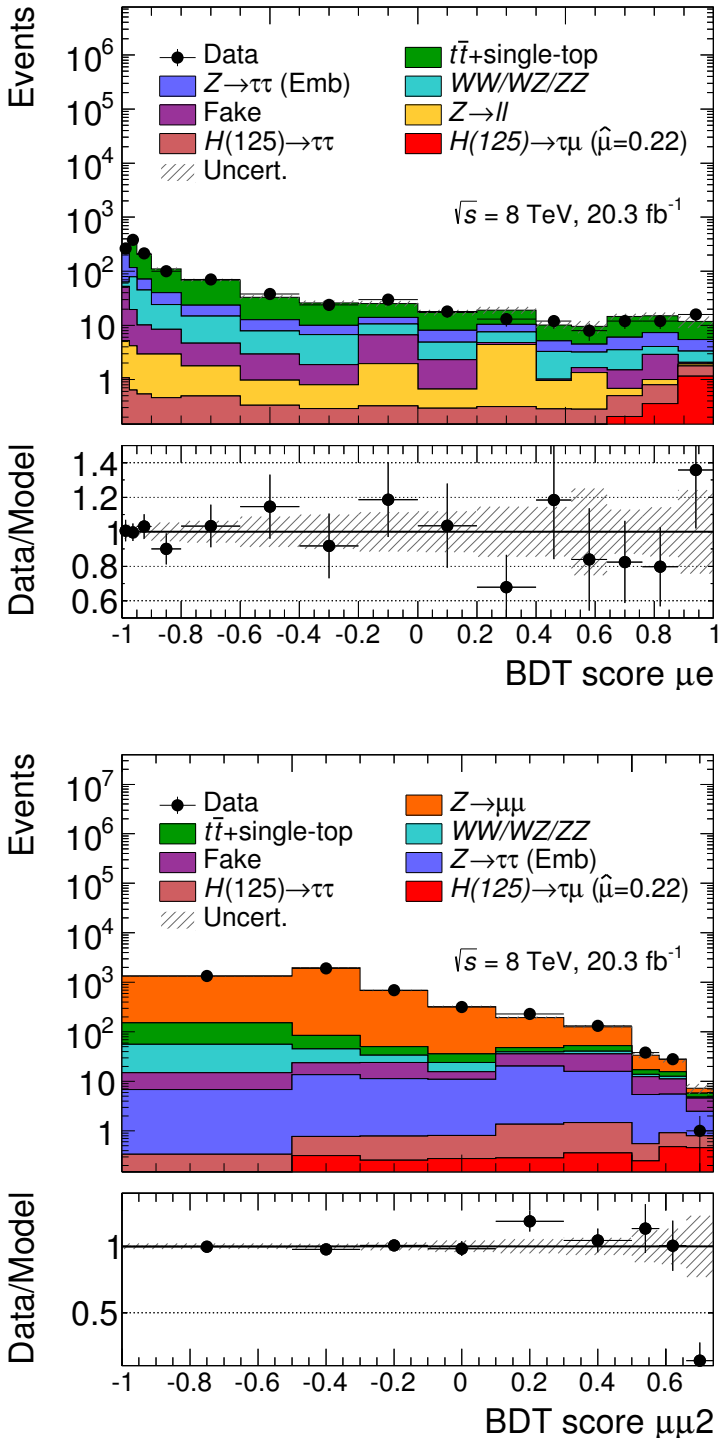


Figure 9.10: Post-fit distribution of the BDT score of the μe (top) and $\mu\mu$ (bottom) signal region for the combined fit in the μe and $\mu\mu$ final state. The uncertainties of both plots correspond to the statistic and systematic uncertainties of signal and background.

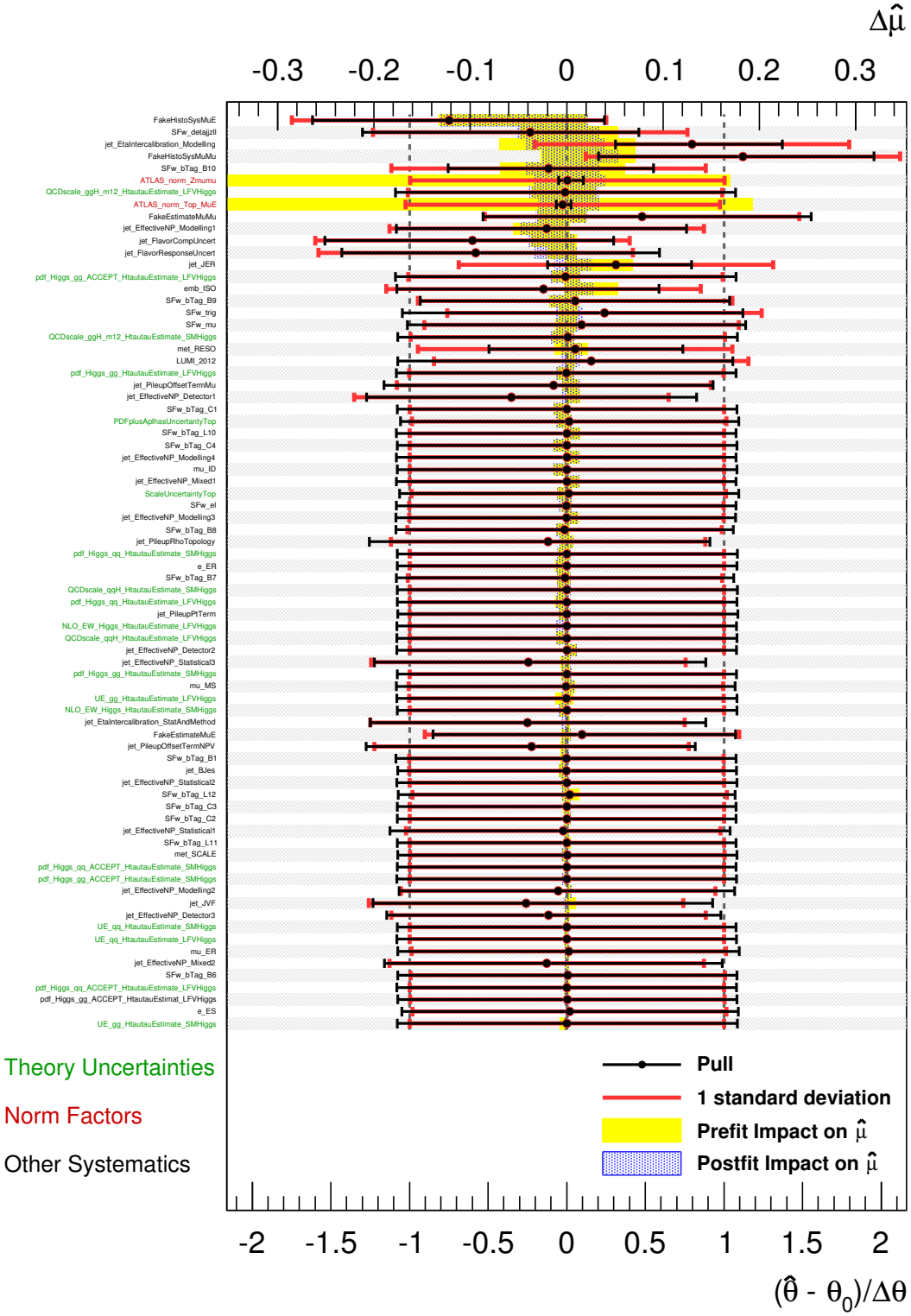


Figure 9.11: Pull-plot for the simultaneous fit for $H \to \tau\mu$ in the $\mu\mu$ and μe final state. A detailed discussion of this pull-plot is given in the main text. Pull-plots in general are explained in the caption of figure 9.2.

9.3.2 Simultaneous Fit for $H \to \tau e$ in the ee and $e\mu$ Final State

This fit for the $H \to \tau e$ signal is performed in all regions of the $e\mu$ and ee final state.

Expected Sensitivity

The fit's ability is tested on Asimov data with a different input value for the signal strength μ. Table 9.24 summarizes these tests by giving the best-fit value $\hat{\mu}$ for the signal strength and the corresponding expected significance. The combined fit is more sensitive than any of the considered final states individually and does not show any bias with respect to the signal strength.

μ in Asimov data	$\hat{\mu}$	expected significance
0	0.00 ± 0.65	0.01
1	1.00 ± 0.77	1.25
5	5.00 ± 1.20	4.88

Table 9.24: Performance of the combined fit for $H \to \tau e$ on Asimov data.

$+2\,\sigma$	2.8 %
$+1\,\sigma$	1.9 %
Median	**1.3** %
$-1\,\sigma$	0.9 %
$-2\,\sigma$	0.7 %

Table 9.25: The upper limit and its uncertainties on the branching ratio for the decay $H \to \tau e$ at a confidence level of 95 % .

The expected upper limit on the branching ratio of the $H \to \tau e$ and its uncertainties are listed in table 9.25. It is with 1.3 % stronger than the limits derived from the individual fits in the $e\mu$ and ee final states of 1.5 % and 3.6 %, respectively.

Results in Data

No hints for the LFV decay $H \to \tau e$ are observed, as can be seen in table 9.26 summarizing the results of the fit performed on data. The best fit signal strength is given by $\hat{\mu} = 0.1 \pm 0.8$ with an observed significance of $0.1\,\sigma$ and is therefore in good agreement with the SM prediction of $\mu = 0$. The observed upper limit is with 1.6 % marginally weaker than the expected limit, but still well within the $1\,\sigma$ confidence interval of the expected limit.

$\hat{\mu}$	observed significance	observed limit BR($H \to \tau\mu$)
0.1 ± 0.8	$0.1\,\sigma$	1.6 %

Table 9.26: Best fit value $\hat{\mu}$ for the signal strength, observed significance and upper limit on the branching ratio (BR) of the decay $H \to \tau e$ at 95 % confidence level.

The pull-plot for this combined fit is given in figure 9.12 showing no significant pulls of any nuisance parameters. The nuisance parameters describing the systematic uncertainties of the fake-lepton estimate are ranked very high meaning that they have a large impact on $\hat{\mu}$. This behavior was

already observed in the fits on the individual final states and stresses that the considered final states might profit in terms of sensitivity from a better fake-lepton background estimate.

Process/Category	$e\mu$ signal region			$e\mu$ Top Quark CR
	Total	2nd last bin	Last bin	Total
Top quark	720 ± 40	7.5 ± 1.0	2.5 ± 0.4	4730 ± 70
$Z \to \tau\tau$ (Emb)	411 ± 20	4.2 ± 0.4	1.76 ± 0.24	99 ± 7
WW/WZ/ZZ	189 ± 6	1.5 ± 0.7	1.14 ± 0.23	24.2 ± 2.4
Fake-lepton	330 ± 40	7.1 ± 1.5	2.9 ± 2.6	117 ± 16
$Z \to ll$	18.1 ± 1.1	0.62 ± 0.08	0.98 ± 0.12	2.9 ± 0.6
VBF $H \to \tau\tau$	4.03 ± 0.22	0.328 ± 0.034	0.35 ± 0.04	0.50 ± 0.04
ggH $H \to \tau\tau$	3.7 ± 1.0	0.17 ± 0.05	0.110 ± 0.032	0.53 ± 0.15
VBF $H \to \tau e$	0 ± 6	< 0.1	0.2 ± 2.5	< 0.1
ggH $H \to \tau e$	0 ± 6	< 0.1	< 0.1	< 0.1
Total Bkd.	1680 ± 40	21.5 ± 2.7	9.7 ± 2.6	4970 ± 70
Data	1684	21	10	4971

Table 9.27: Expected event yield and data in the $e\mu$ signal region and in the $e\mu$ top quark control region (CR) after a simultaneous fit in the $e\mu$ and ee final state. The composition of the last and 2nd last bin of the BDT score (two highest bins in the BDT score) of the $e\mu$ signal region is given additionally. The uncertainties of the individual processes are only systematic uncertainties, while the given uncertainty on the total background (Bkd.) corresponds to systematic and statistical uncertainties.

Process/Category	ee Signal Region			$Z \to ee$ CR
	Total	2nd last bin	Last bin	Total
$Z \to ee$	3050 ± 60	22.3 ± 2.4	1.70 ± 0.22	60540 ± 270
Top quark	122 ± 11	5.3 ± 0.6	1.78 ± 0.33	340 ± 29
$Z \to \tau\tau$ (Emb)	28.1 ± 2.7	2.17 ± 0.27	1.10 ± 0.18	84 ± 5
WW/WZ/ZZ	57.2 ± 2.8	0.64 ± 0.08	0.52 ± 0.08	432 ± 11
Fake-lepton	48 ± 23	4.0 ± 1.2	0.7 ± 1.4	140 ± 70
VBF $H \to \tau\tau$	1.63 ± 0.10	0.184 ± 0.018	0.259 ± 0.034	0.86 ± 0.05
ggH $H \to \tau\tau$	0.85 ± 0.23	< 0.1	0.14 ± 0.04	1.09 ± 0.29
VBF $H \to \tau e$	0.2 ± 3.1	< 0.1	0.1 ± 0.8	0.1 ± 0.9
ggH $H \to \tau e$	0.2 ± 2.1	< 0.1	< 0.1	0.2 ± 2.0
Total Bkd.	3310 ± 60	34.7 ± 3.3	6.2 ± 1.7	61540 ± 260
Data	3309	29	6	61544

Table 9.28: Expected event yield and data in the ee signal region and in the $Z \to ee$ control region (CR) after a simultaneous fit in the $e\mu$ and ee final state. The composition of the last and 2nd last bin of the BDT score (two highest bins in the BDT score) of the ee signal region is given additionally. The uncertainties of the individual processes are only systematic uncertainties, while the given uncertainty on the total background (Bkd.) corresponds to systematic and statistical uncertainties.

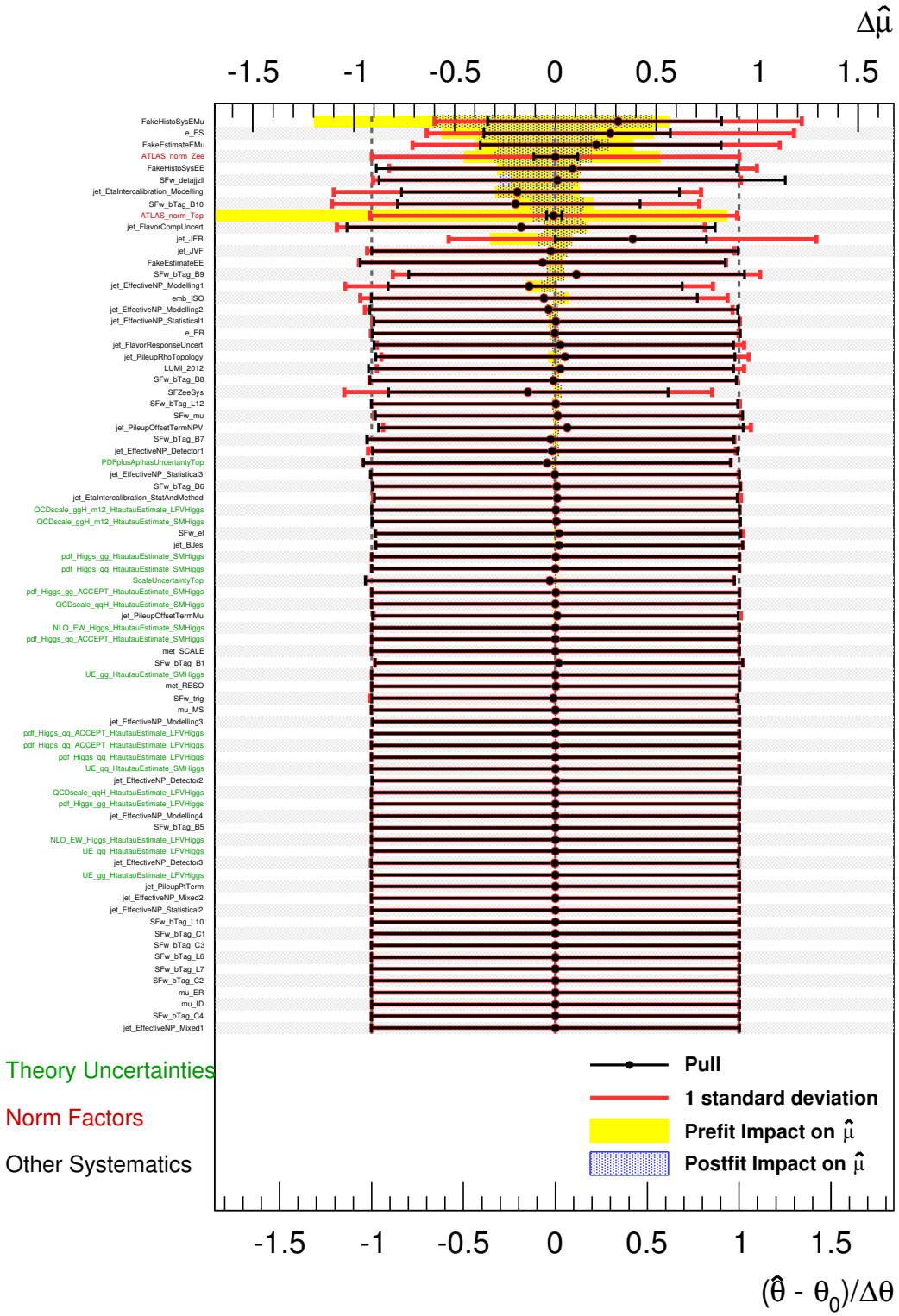


Figure 9.12: Pull-plot for the combined fit for $H \to \tau e$ in the ee and $e\mu$ final state. A detailed discussion of this pull-plot is given in the main text. Pull-plots in general are explained in the caption of figure 9.2.

The post-fit event yields in for all signal and control regions used by the combined fit are given in table 9.27 and 9.28. They show no significant deviation between observed data and the the sum of fitted signal and post-fit background estimates. The same holds true for the post-fit plots of the distribution of the BDT score in the $e\mu$ and ee signal region and the composition of the $Z \to ee$ and $e\mu$ top quark control region shown in figure 9.13 and 9.14, respectively. The combination of the good agreement of observed data and post-fit background and signal estimates and the well behaved pull-plot make the results of this combined fit very trustworthy.

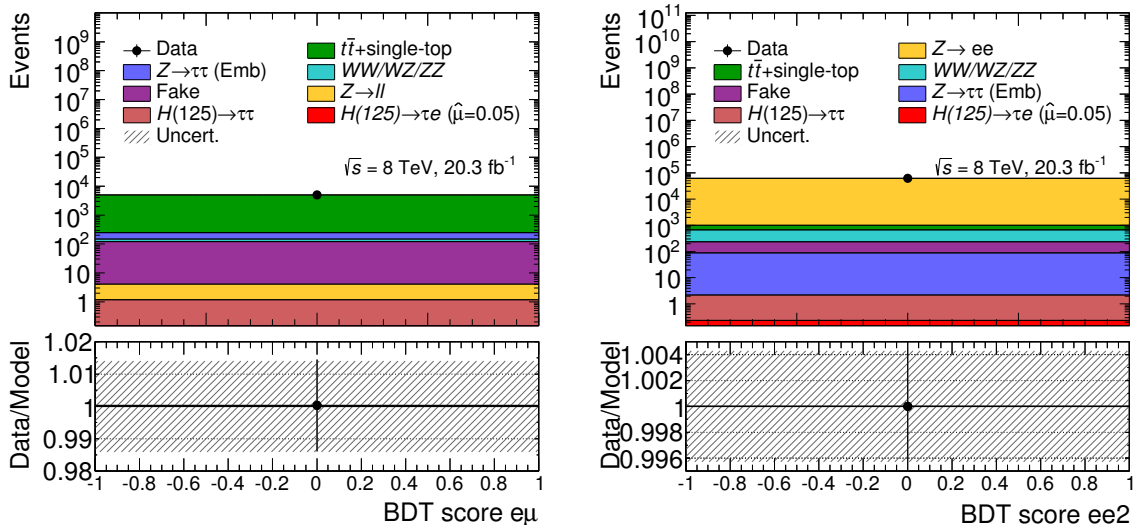


Figure 9.13: Post-fit composition of the $e\mu$ top quark control region (left) and the $Z \to ee$ control region (right) using the best-fit parameters from the simultaneous fit in the $e\mu$ and ee final state.

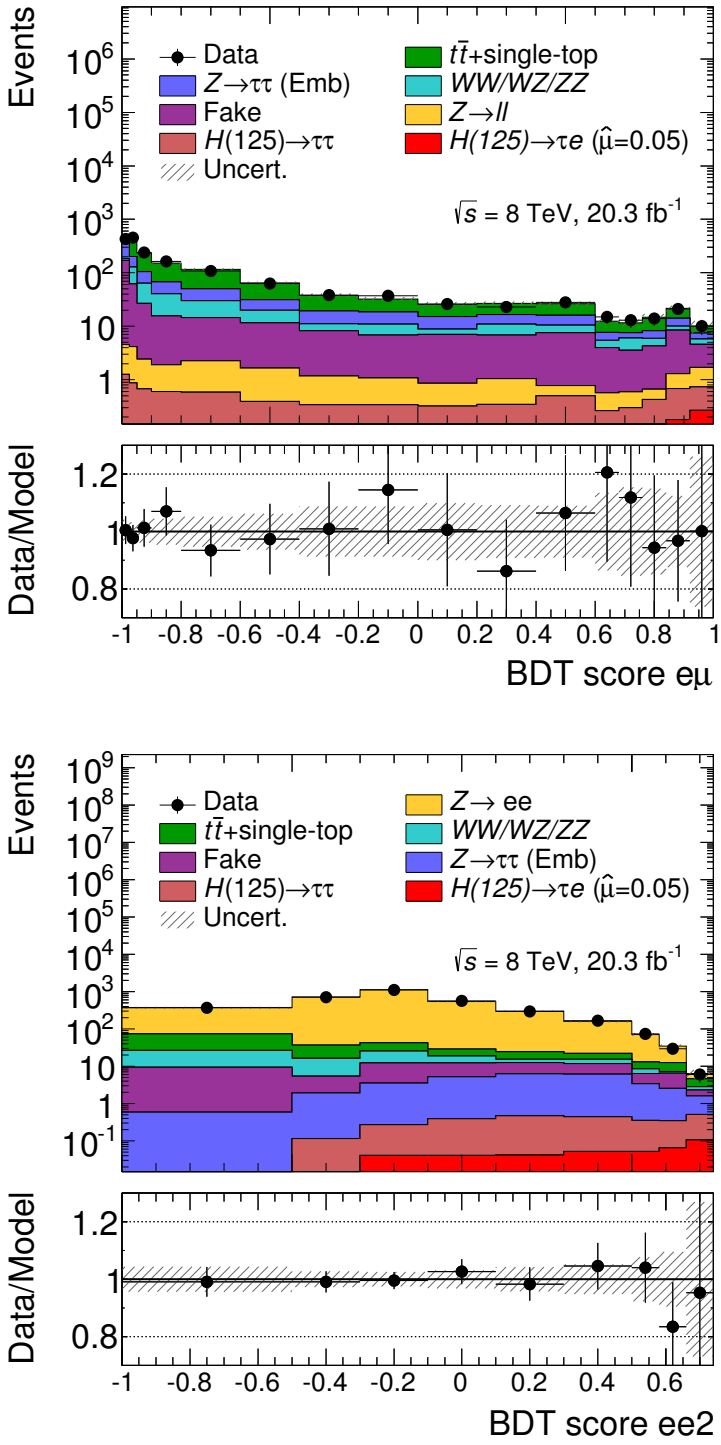


Figure 9.14: Post-fit distributions of the BDT score in the $e\mu$ signal region (top) and the ee signal region (bottom) using the best fit parameters from the combined fit for $H \to \tau e$ in the ee and $e\mu$ final state.

9.4 Summary of the Results

The results of the fits in the individual final states and the combined fits are summarized in table 9.29. All fits of the individual final states are in good agreement with the combined fit,

	Results for $H \to \tau\mu$ in			**Results for $H \to \tau e$ in**		
	μe	$\mu\mu$	$\boldsymbol{\mu e + \mu\mu}$	$e\mu$	ee	$\boldsymbol{e\mu + ee}$
Exp. upper limit	$1.7^{+0.9}_{-0.5}\,\%$	$2.7^{+1.5}_{-0.8}\,\%$	$\mathbf{1.4^{+0.7}_{-0.4}\,\%}$	$1.5^{+0.7}_{-0.5}\,\%$	$3.6^{+2.4}_{-1.2}\,\%$	$\mathbf{1.3^{+0.6}_{-0.4}\,\%}$
Obs. upper limit	$2.5\,\%$	$1.5\,\%$	$\mathbf{1.5\,\%}$	$1.9\,\%$	$3.3\,\%$	$\mathbf{1.6\,\%}$
$\hat{\mu}$	0.9 ± 0.8	-3.5 ± 1.0	$\mathbf{0.2 \pm 0.6}$	0.1 ± 1.0	-0.5 ± 1.7	$\mathbf{0.1 \pm 0.8}$
Obs. Significance	$1.2\,\sigma$	$-2.6\,\sigma$	$\mathbf{0.4\,\sigma}$	$0.1\,\sigma$	$-0.3\,\sigma$	$\mathbf{0.1\,\sigma}$

Table 9.29: Summary of the results of the fit in each final state and the combined fits. The stated expected (Exp.) and observed (Obs.) upper limits on the branching ratio of the considered signal are calculated at 95 % confidence level. The best fit value $\hat{\mu}$ of the signal strength μ and the observed significance of this fitted signal is given as well. A signal of $\mu = 1$ corresponds to a branching ratio of 1 % for the corresponding decay.

except the fit of the $\mu\mu$ final state. In this final state, an overestimation of the background in the highest bin of the BDT score compared to the observed data yields a negative fitted signal strength that is not compatible with $\mu = 0$. The observed significance of this lack of data compared to the background estimates is $-2.6\,\sigma$. A further investigation of this issue, that was discussed in more detail in section 9.2.2, was not possible within the timescale of this master's thesis.
Furthermore, it can be seen that the different flavor final states are significantly more sensitive than the same flavor final states as they yield much stricter expected limits. The same flavor final states allow nevertheless to derive stricter limits in the combined fits. The improvement of the expected limits, due to the combination, is however relatively small. This means that it would be in principle desirable not to refrain from the same flavor final states in a future search for LFV to be able to derive as strict limits as possible. However, it should be considered carefully whether the same amount of effort that is necessary to carry out the analysis in the same flavor final states being put in an expansion of the analysis to the different flavor final states beyond the VBF topology would yield more sensitive results.

Chapter 10

Conclusion

A search for the LFV Higgs boson decays $H \to \tau\mu$ and $H \to \tau e$ of the recently discovered Higgs boson at a mass of 125.09 GeV has been performed. The analysis of the presented search uses the topology of the VBF production mode and considers all fully leptonic final states. It was carried out on the full data set recorded by the ATLAS detector at a center-of-mass energy $\sqrt{s} = 8\,\text{TeV}$ corresponding to an integrated luminosity $20.3\,\text{fb}^{-1}$.

The new concept of lepton labeling is introduced. It allows to exploit the specific and distinguished properties of the two detected leptons of the LFV Higgs boson decay caused by their different parentage: one of the leptons originates directly from the Higgs boson decay, while the other one originates from the decay of the τ lepton. The used lepton labeling procedure allows to identify the leptons origin with an efficiency of $\mathcal{O}(90\,\%)$ and assigns the two detected leptons the labels l_H and l_τ according to their reconstructed origin.

The analysis is split into four final states to utilize the different background and signal contributions. Two same flavor final states, ee and $\mu\mu$, are formed. They are dominated by the overwhelming $Z \to ee$ and $Z \to \mu\mu$ backgrounds, respectively. The events containing two leptons of different flavor are split in the μe and $e\mu$ final state depending on the flavor of the lepton labeled as l_H. Events are sorted into the μe ($e\mu$) final state if l_H is a muon (electron). Both same flavor final states are dominated by the top quark background. The $\mu\mu$ and μe (ee and $e\mu$) final state are enriched in the $H \to \tau\mu$ ($H \to \tau e$) signal and their contamination with the *wrong* signal, meaning the $H \to \tau e$ ($H \to \tau\mu$) signal, was found to be negligible. Each final state is split into a signal region and a control region for the dominant background. These control regions are used to infer the normalization of the dominant background for this final state from data.

The main challenge in the search for LFV decays of the Higgs boson is dealing with a tiny a-priori signal to background ratio and a potentially overall small signal yield. The presented analysis is therefore based on the multivariate analysis tool of Boosted Decision Trees (BDT) that provide a good separation between signal and background, while sparing the rare signal. The output of the BDTs, the so-called BDT score, separates the signal well from all background processes and is hence used as the final discriminant.

Individual BDTs are trained in the signal region of each final state. Then, the distributions of their BDT score in the signal region and the event yields of the control regions are fitted to extract information on the signal processes in the statistical evaluation of the analysis. The used fits consider all dominant systematic uncertainties. First, a fit is performed in each of the final states separately, providing two independent measurements of both considered LFV decays. Then, a combined fit for the $H \to \tau\mu$ ($H \to \tau e$) signal is performed simultaneously in the μe and $\mu\mu$ ($e\mu$ and ee) final states.

The best fit value of the combined fits for the $H \to \tau\mu$ and $H \to \tau e$ are $\hat{\mu} = 0.2 \pm 0.6$ and $\hat{\mu} = 0.1 \pm 0.8$, respectively. A signal strength of $\mu = 1$ corresponds to a branching ratio of 1 %. Both best fit values of the signal strengths are in good agreement with the SM prediction of $\mu = 0$. The observed (expected) upper limits at 95 % confidence level on the branching ratio of $H \to \tau\mu$ and $H \to \tau e$ are 1.5 % (1.4 %) and 1.6 % (1.3 %), respectively. These observed limits are by an order of magnitude better than the indirect limits. The main contribution to the sensitivity of the combined fits originates from the different flavor final states.

Both, the CMS and the ATLAS collaborations have published searches for LFV decay $H \to \tau\mu$ of a Higgs boson with a mass of 125 GeV. The CMS collaboration has performed a search for the decay $H \to \tau\mu$ [35] using the decay channels $H \to \tau\mu \to \tau_{\text{had}}e$ and $H \to \tau\mu \to e\mu$, where τ_{had} denotes a hadronic decay of the τ lepton. Their best fit-value for μ is $\hat{\mu} = 0.84^{+0.39}_{-0.37}$ corresponding to a small excess with a significance of $2.4\,\sigma$. The observed (expected) limit on the branching ratio of the $H \to \tau\mu$ is 1.51 % (0.75 %) at 95 % confidence level. The ATLAS collaboration has published a search for the LFV decay $H \to \tau\mu$ [36] in the decay channel $H \to \tau\mu \to \tau_{\text{had}}\mu$. This analysis yields an observed (expected) upper limit on the branching ratio of the decay $H \to \tau\mu$ of 1.85 % (1.25 %) and a best fit signal strength of $\hat{\mu} = 0.77 \pm 0.62$. The results derived in the analysis presented in this thesis for the $H \to \tau\mu$ are in good agreement with the published results from the ATLAS and CMS collaboration, since their best fit value for the signal strength is covered roughly by the $1\,\sigma$ uncertainties of the best fit value derived in this analysis. The observed upper limits on the branching ratio of $H \to \tau\mu$ of this analysis is compatible with the limits observed by CMS and ATLAS.

So far, only the CMS collaboration has published results on the search for the lepton flavor violating decay $H \to \tau e$ [118]. Their search considers the decay channels $H \to \tau e \to \tau_{\text{had}}e$ and $H \to \tau e \to \mu e$ and did not show any excess. Their best fit value for the signal strength is consequently with $\hat{\mu} = -0.1^{+0.37}_{-0.36}$ in good agreement with the SM prediction of $\mu = 0$. They set preliminary observed (expected) limits on the branching ratio of the $H \to \tau e$ decay of 0.69 % ($0.75^{+0.32}_{-0.22}$ %). The best fit values of the signal strength for the $H \to \tau e$ decay of the presented and the CMS analysis are in good agreement, while the CMS experiment is able to set a stricter limit, which is not surprising since the CMS analysis considers more decay modes than the presented analysis. Their observed (expected) limit in the fully leptonic decay channel with two jets, that is to some extent compatible to the presented analysis, is 1.49 % ($1.59^{+0.93}_{-0.55}$ %) yielding about the same sensitivity for both analyses in this decay channel.

There is currently an effort in the ATLAS collaboration to publish results using the $H \to \tau e$ decay and to expand the analysis of the $H \to \tau\mu$ decay to the different flavor fully leptonic final state, but none of these results were published at the time this thesis was written.

The presented analysis could be improved, first of all, by investigating and removing the potential miss-modeling of the BDT score in the $\mu\mu$ final state and by improving the estimate of the fake-lepton background to minimize its influence on the fitted signal strength. Furthermore, the presented analysis makes only use of the fully leptonic final states in the VBF topology. Expanding the analysis to fully leptonic final states beyond the VBF topology and combining it with an analysis of the semi-leptonic final states ($H \to \tau_{\text{had}}\mu$ and $H \to \tau_{\text{had}}e$) will most likely help to improve the sensitivity of the analysis significantly. This and the additional dataset recorded during the Run 2 of the LHC would help to investigate whether LFV decays of the Higgs boson are realized in nature.

Bibliography

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

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

[3] Peter W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964.

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

[5] Peter W. Higgs. Spontaneous Symmetry Breakdown without Massless Bosons. *Phys. Rev.*, 145:1156–1163, May 1966.

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

[7] S. L. Glashow. Partial Symmetries of Weak Interactions. *Nucl. Phys.*, 22:579–588, 1961.

[8] Steven Weinberg. A model of leptons. *Phys. Rev. Lett.*, 19:1264–1266, Nov 1967.

[9] A. Salam. Elementary particle theory. *Ed. N. Svartholm, Stockholm, Almquist and Wiksell*, 367, 1968.

[10] ATLAS Collaboration. 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(1):1 – 29, 2012.

[11] CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters B*, 716(1):30 – 61, 2012.

[12] Super-Kamiokande Collaboration. Evidence for Oscillation of Atmospheric Neutrinos. *Phys. Rev. Lett.*, 81:1562–1567, Aug 1998.

[13] Super-Kamiokande Collaboration. Measurement of Atmospheric Neutrino Flux Consistent with Tau Neutrino Appearance. *Phys. Rev. Lett.*, 97:171801, Oct 2006.

[14] Super-Kamiokande Collaboration. Atmospheric neutrino oscillation analysis with subleading effects in Super-Kamiokande I, II, and III. *Phys. Rev. D*, 81:092004, May 2010.

[15] SNO Collaboration. Measurement of the Rate of $\nu_e + d \rightarrow p + p + e^-$ Interactions Produced by ^8B Solar Neutrinos at the Sudbury Neutrino Observatory. *Phys. Rev. Lett.*, 87:071301, Jul 2001.

[16] B. Pontecorvo. Mesonium and anti-mesonium. *Sov. Phys. JETP*, 6:429, 1957. [Zh. Eksp. Teor. Fiz.33,549(1957)].

[17] B. Pontecorvo. Inverse beta processes and nonconservation of lepton charge. *Sov. Phys. JETP*, 7:172–173, 1958. [Zh. Eksp. Teor. Fiz.34,247(1957)].

[18] Ziro Maki, Masami Nakagawa, and Shoichi Sakata. Remarks on the Unified Model of Elementary Particles. *Progress of Theoretical Physics*, 28(5):870–880, 1962.

[19] James D. Bjorken and Steven Weinberg. Mechanism for Nonconservation of Muon Number. *Phys. Rev. Lett.*, 38:622–625, Mar 1977.

[20] J. Lorenzo Diaz-Cruz and J. J. Toscano. Lepton flavor violating decays of Higgs bosons beyond the standard model. *Phys. Rev. D*, 62:116005, Nov 2000.

[21] Tao Han and Danny Marfatia. $h \to \mu\tau$ at Hadron Colliders. *Phys. Rev. Lett.*, 86:1442–1445, Feb 2001.

[22] Abdesslam Arhrib, Yifan Cheng, and Otto C. W. Kong. Comprehensive analysis on lepton flavor violating Higgs boson to $\mu^{\mp}\tau^{\pm}$ decay in supersymmetry without R parity. *Phys. Rev. D*, 87:015025, Jan 2013.

[23] Kaustubh Agashe and Roberto Contino. Composite Higgs-mediated flavor-changing neutral current. *Phys. Rev. D*, 80:075016, Oct 2009.

[24] Aleksandr Azatov, Manuel Toharia, and Lijun Zhu. Higgs mediated flavor changing neutral currents in warped extra dimensions. *Phys. Rev. D*, 80:035016, Aug 2009.

[25] Hajime Ishimori et al. Non-Abelian Discrete Symmetries in Particle Physics. *Progress of Theoretical Physics Supplement*, 183:1–163, 2010.

[26] Gilad Perez and Lisa Randall. Natural neutrino masses and mixings from warped geometry. *Journal of High Energy Physics*, 2009(01):077, 2009.

[27] Monika Blanke, Andrzej J. Buras, Björn Duling, Stefania Gori, and Andreas Weiler. $\Delta F = 2$ observables and fine-tuning in a warped extra dimension with custodial protection. *Journal of High Energy Physics*, 2009(03):001, 2009.

[28] Gian F. Giudice and Oleg Lebedev. Higgs-dependent Yukawa couplings . *Physics Letters B*, 665(23):79 – 85, 2008.

[29] J.A. Aguilar-Saavedra. A minimal set of top-Higgs anomalous couplings. *Nuclear Physics B*, 821(12):215 – 227, 2009.

[30] Michaela E. Albrecht, Monika Blanke, Andrzej J. Buras, Bjrn Duling, and Katrin Gemmler. Electroweak and flavour structure of a warped extra dimension with custodial protection. *Journal of High Energy Physics*, 2009(09):064, 2009.

[31] Andreas Goudelis and Oleg Lebedev and Jae-hyeon Park. Higgs-induced lepton flavor violation. *Physics Letters B*, 707(34):369 – 374, 2012.

[32] David McKeen, Maxim Pospelov, and Adam Ritz. Modified Higgs branching ratios versus CP and lepton flavor violation. *Phys. Rev. D*, 86:113004, Dec 2012.

[33] Gianluca Blankenburg, John Ellis, and Gino Isidori. Flavour-changing decays of a 125 GeV Higgs-like particle . *Physics Letters B*, 712(45):386 – 390, 2012.

[34] Roni Harnik, Joachim Kopp, and Jure Zupan. Flavor violating higgs decays. *Journal of High Energy Physics*, 2013(3), 2013.

[35] CMS Collaboration. Search for lepton-flavour-violating decays of the Higgs boson. *Physics Letters B*, 749:337 – 362, 2015.

[36] ATLAS Collaboration. Search for lepton-flavour-violating $H \to \mu\tau$ decays of the Higgs boson with the ATLAS detector. 2015.

[37] ATLAS Collaboration. Evidence for the Higgs-boson Yukawa coupling to tau leptons with the ATLAS detector. *Journal of High Energy Physics*, 2015(4), 2015.

[38] K. A. Olive et al. Review of Particle Physics. *Chin. Phys.*, C38:090001, 2014.

[39] S. Dittmaier and M. Schumacher. The Higgs Boson in the Standard Model - From LEP to LHC: Expectations, Searches, and Discovery of a Candidate. *Prog. Part. Nucl. Phys.*, 70:1–54, 2013.

[40] Y. Nambu and G. Jona-Lasinio. Dynamical Model of Elementary Particles Based on an Analogy with Superconductivity. I. *Phys. Rev.*, 122:345–358, Apr 1961.

[41] M. Gell-Mann and M. Lvy. The axial vector current in beta decay. *Il Nuovo Cimento (1955-1965)*, 16(4):705–726, 1960.

[42] Nicola Cabibbo. Unitary Symmetry and Leptonic Decays. *Phys. Rev. Lett.*, 10:531–533, Jun 1963.

[43] Makoto Kobayashi and Toshihide Maskawa. CP-Violation in the Renormalizable Theory of Weak Interaction. *Progress of Theoretical Physics*, 49(2):652–657, 1973.

[44] J. Adam et al. New constraint on the existence of the $\mu^+ \to e^+\gamma$ decay. *Phys. Rev. Lett.*, 110:201801, May 2013.

[45] K. Hayasaka et al. New search for $\tau \to \mu\gamma$ and $\tau \to e\gamma$ decays at Belle. *Physics Letters B*, 666(1):16 – 22, 2008.

[46] B. Aubert et al. Searches for lepton flavor violation in the decays $\tau^\pm \to e^\pm\gamma$ and $\tau^\pm \to \mu^\pm\gamma$. *Phys. Rev. Lett.*, 104:021802, Jan 2010.

[47] LHC Higgs Cross Section Working Group, S. Heinemeyer, C. Mariotti, G. Passarino, and R. Tanaka (Eds.). Handbook of LHC Higgs Cross Sections: 3. Higgs Properties. *CERN-2013-004*, CERN, Geneva, 2013. arXiv:1307.1347.

[48] `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CrossSections`. 28^{th} of October.

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

[50] The ATLAS and CMS Collaborations. Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV. Technical Report ATLAS-CONF-2015-044, CERN, Geneva, Sep 2015.

[51] ATLAS Collaboration. Evidence for the spin-0 nature of the Higgs boson using ATLAS data. *Physics Letters B*, 726(13):120 – 144, 2013.

[52] CMS Collaboration. Properties of the Higgs-like boson in the decay H to ZZ to 4l in pp collisions at sqrt s =7 and 8 TeV. Technical Report CMS-PAS-HIG-13-002, CERN, Geneva, 2013.

[53] ATLAS Collaboration. Measurements of the Higgs boson production and decay rates and coupling strengths using *pp* collision data at $\sqrt{s} = 7$ and 8 TeV in the ATLAS experiment. 2015. arXiv:1507.04548 [hep-ex].

[54] CMS Collaboration. Precise determination of the mass of the higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV. *The European Physical Journal C*, 75(5), 2015.

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

[56] ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *Journal of Instrumentation*, 3(08):S08003, 2008.

[57] Lyndon Evans and Philip Bryant. LHC Machine. *Journal of Instrumentation*, 3(08):S08001, 2008.

[58] ATLAS Collaboration. Performance of the ATLAS Trigger System in 2010. *The European Physical Journal C*, 72(1), 2012.

[59] Paolo Nason. A new method for combining nlo qcd with shower monte carlo algorithms. *Journal of High Energy Physics*, 2004(11):040, 2004.

[60] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching nlo qcd computations with parton shower simulations: the powheg method. *Journal of High Energy Physics*, 2007(11):070, 2007.

[61] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing nlo calculations in shower monte carlo programs: the powheg box. *Journal of High Energy Physics*, 2010(6), 2010.

[62] E. Bagnaschi, G. Degrassi, P. Slavich, and A. Vicini. Higgs production via gluon fusion in the powheg approach in the sm and in the mssm. *Journal of High Energy Physics*, 2012(2), 2012.

[63] Torbjörn Sjöstrand and Stephen Mrenna and Peter Skands. A brief introduction to PYTHIA 8.1. *Computer Physics Communications*, 178(11):852 – 867, 2008.

[64] Summary of ATLAS Pythia 8 tunes. ATL-PHYS-PUB-2012-003, 2012.

[65] H.-L. Lai et al. New generation of parton distributions with uncertainties from global QCD analysis. *Phys. Rev.*, D 82:074024, 2010.

[66] D. J. Lange. The EvtGen particle decay simulation package. *Nucl. Instrum. Meth.*, A462:152–155, 2001.

[67] Michelangelo L. Mangano, Fulvio Piccinini, Antonio D. Polosa, Mauro Moretti, and Roberto Pittau. Alpgen, a generator for hard multiparton processes in hadronic collisions. *Journal of High Energy Physics*, 2003(07):001, 2003.

[68] Stefano Catani, Leandro Cieri, Giancarlo Ferrera, Daniel de Florian, and Massimiliano Grazzini. Vector Boson Production at Hadron Colliders: A Fully Exclusive QCD Calculation at Next-to-Next-to-Leading Order. *Phys. Rev. Lett.*, 103:082001, Aug 2009.

[69] Stefano Catani and Massimiliano Grazzini. Next-to-next-to-leading-order subtraction formalism in hadron collisions and its application to higgs-boson production at the large hadron collider. *Phys. Rev. Lett.*, 98:222002, May 2007.

[70] T. Gleisberg and S. Höche and F. Krauss and M. Schönherr and S. Schumann and F. Siegert and J. Winter. Event generation with SHERPA 1.1. *Journal of High Energy Physics*, 2009(02):007, 2009.

[71] Nikolaos Kidonakis. NNLL resummation for s-channel single top quark production. *Phys.Rev.*, D81:054028, 2010.

[72] Borut Paul Kersevan and Elzbieta Richter-Was. The Monte Carlo event generator AcerMC versions 2.0 to 3.8 with interfaces to PYTHIA 6.4, HERWIG 6.5 and ARIADNE 4.1. *Computer Physics Communications*, 184(3):919 – 985, 2013.

[73] Torbjörn Sjöstrand and Stephen Mrenna and Peter Skands. PYTHIA 6.4 physics and manual. *Journal of High Energy Physics*, 2006(05):026, 2006.

[74] Nikolaos Kidonakis. Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production. *Phys.Rev.*, D83:091503, 2011.

[75] Nikolaos Kidonakis. Two-loop soft anomalous dimensions for single top quark associated production with a W- or H-. *Phys.Rev.*, D82:054018, 2010.

[76] Matteo Cacciari, Michal Czakon, Michelangelo Mangano, Alexander Mitov, and Paolo Nason. Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation. *Phys.Lett.*, B710:612–622, 2012.

[77] Peter Brnreuther, Michal Czakon, and Alexander Mitov. Percent Level Precision Physics at the Tevatron: First Genuine NNLO QCD Corrections to $q\bar{q} \to t\bar{t} + X$. *Phys.Rev.Lett.*, 109:132001, 2012.

[78] Michal Czakon and Alexander Mitov. NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels. *JHEP*, 1212:054, 2012.

[79] Michal Czakon and Alexander Mitov. NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction. *JHEP*, 1301:080, 2013.

[80] Micha Czakon, Paul Fiedler, and Alexander Mitov. Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_S^4)$. *Phys.Rev.Lett.*, 110:252004, 2013.

[81] Michal Czakon and Alexander Mitov. Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders. *Comput.Phys.Commun.*, 185:2930, 2014.

[82] G. Corcella et al. Herwig 6.5 release note. *arXiv:hep-ph/0210213*, 2002.

[83] John M. Campbell, R. Keith Ellis, and Ciaran Williams. Vector boson pair production at the LHC. *JHEP*, 1107:018, 2011.

[84] Thomas Binoth and Mariano Ciccolini and Nikolas Kauer and Michael Krämer. Gluon-induced W-boson pair production at the LHC. *Journal of High Energy Physics*, 2006(12):046, 2006.

[85] A. Djouadi, M. Spira, and P.M. Zerwas. Production of Higgs bosons in proton colliders. QCD corrections. *Physics Letters B*, 264(3-4):440–446, 1991.

[86] S. Dawson. Radiative corrections to Higgs boson production . *Nuclear Physics B*, 359(2-3):283–300, 1991.

[87] M. Spira, A. Djouadi, D. Graudenz, and R.M. Zerwas. Higgs boson production at the LHC. *Nuclear Physics B*, 453(1-2):17–82, 1995.

[88] Robert V. Harlander and William B. Kilgore. Next-to-Next-to-Leading Order Higgs Production at Hadron Colliders. *Phys. Rev. Lett.*, 88:201801, May 2002.

[89] Charalampos Anastasiou and Kirill Melnikov. Higgs boson production at hadron colliders in NNLO QCD. *Nuclear Physics B*, 646(1-2):220–256, 2002.

[90] V. Ravindran, J. Smith, and W.L. van Neerven. NNLO corrections to the total cross section for Higgs boson production in hadron-hadron collisions. *Nuclear Physics B*, 665(0):325–366, 2003.

[91] Stefano Catani et al. Soft-gluon resummation for higgs boson production at hadron colliders. *JHEP*, 07(028), 2003.

[92] U. Aglietti, R. Bonciani, G. Degrassi, and A. Vicini. Two-loop light fermion contribution to higgs production and decays. *Physics Letters B*, 595(1-4):432 – 441, 2004.

[93] Stefano Actis, Giampiero Passarino, Christian Sturm, and Sandro Uccirati. NLO electroweak corrections to Higgs boson production at hadron colliders . *Physics Letters B*, 670(1):12–17, 2008.

[94] M. Ciccolini, A. Denner, and S. Dittmaier. Strong and Electroweak Corrections to the Production of a Higgs Boson + 2 Jets via Weak Interactions at the Large Hadron Collider. *Phys. Rev. Lett.*, 99:161803, Oct 2007.

[95] M. Ciccolini, Denner A., and Dittmaier S. Electroweak and QCD corrections to Higgs production via vector-boson fusion at the CERN LHC. *Phys. Rev. D*, 77:013002, Jan 2008.

[96] K. Arnold et al. Vbfnlo: A parton level Monte Carlo for processes with electroweak bosons . *Computer Physics Communications*, 180(9):1661–1670, 2009.

[97] ATLAS Collaboration. The ATLAS Simulation Infrastructure. *Eur. Phys. J.*, C 70:823, 2010.

[98] GEANT4 Collaboration, S. Agostinelli et al. Geant4– A Simulation toolkit. *Nucl. Instrum. Meth.*, A 506:250, 2003.

[99] ATLAS Collaboration. Modelling $Z \to \tau\tau$ processes in ATLAS with τ-embedded $Z \to \mu\mu$ data. *JINST*, 10(09):P09018, 2015.

[100] ATLAS Collaboration. Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC protonproton collision data. *The European Physical Journal C*, 74(11), 2014.

[101] ATLAS Collaboration. Electron reconstruction and identification efficiency measurements with the ATLAS detector using the 2011 LHC protonproton collision data. *The European Physical Journal C*, 74(7), 2014.

[102] Electron efficiency measurements with the ATLAS detector using the 2012 LHC protonproton collision data. Technical Report ATLAS-CONF-2014-032, CERN, Geneva, Jun 2014.

[103] Matteo Cacciari and Gavin P. Salam. Dispelling the myth for the jet-finder. *Physics Letters B* , 641(1):57 – 61, 2006.

[104] Matteo Cacciari and Gavin P. Salam and Gregory Soyez. The anti-k$_t$ jet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063, 2008.

[105] W. Lampl, S. Laplace, D. Lelas, P. Loch, H. Ma, S. Menke, S. Rajagopalan, D. Rousseau, S. Snyder, and G. Unal. Calorimeter Clustering Algorithms: Description and Performance. Technical Report ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003, CERN, Geneva, Apr 2008.

[106] ATLAS Collaboration. Performance of Missing Transverse Momentum Reconstruction in ATLAS studied in Proton-Proton Collisions recorded in 2012 at 8 TeV. Technical Report ATLAS-CONF-2013-082, CERN, Geneva, Aug 2013.

[107] L. Breiman, J. Friedman, R. Olshen, and C. Stone. *Classification and Regression Trees*. Chapman & Hall, 1984.

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

[109] Yoav Freund and Robert E Schapire. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*, 55(1):119 – 139, 1997.

[110] Andreas Hoecker, Peter Speckmayer, Joerg Stelzer, Jan Therhaag, Eckhard von Toerne, and Helge Voss. TMVA: Toolkit for Multivariate Data Analysis. *PoS*, ACAT:040, 2007. arXiv:physics/0703039.

[111] Gini, C. Variabilita e mutabilita. *Memorie di metodologia statistica*, 1912.

[112] R.K. Ellis, I. Hinchliffe, M. Soldate, and J.J. Van Der Bij. Higgs decay to $\tau^+\tau^-$ A possible signature of intermediate mass Higgs bosons at high energy hadron colliders. *Nuclear Physics B*, 297(2):221 – 243, 1988.

[113] ATLAS Collaboration. Calibration of the performance of b-tagging for c and light-flavour jets in the 2012 ATLAS data. Technical Report ATLAS-CONF-2014-046, CERN, Geneva, July 2014.

[114] ATLAS Collaboration. Combination of ATLAS and CMS top quark pair cross section measurements in the $e\mu$ final state using proton-proton collisions at $\sqrt{s}$= 8 TeV. Technical Report ATLAS-CONF-2014-054, CERN, Geneva, Sep 2014.

[115] Roger Barlow. Extended maximum likelihood. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 297(3):496 – 506, 1990.

[116] Kyle Cranmer, George Lewis, Lorenzo Moneta, Akira Shibata, and Wouter Verkerke. HistFactory: A tool for creating statistical models for use with RooFit and RooStats. Technical Report CERN-OPEN-2012-016, New York U., New York, Jan 2012.

[117] A. L. Read. Presentation of search results: the CL_s technique. *Journal of Physics G: Nuclear and Particle Physics*, 28(10):2693, 2002.

[118] CMS Collaboration. Search for lepton-flavour-violating decays of the Higgs boson to $e\tau$ and $e\mu$ at $\sqrt{s} = 8$ TeV. Technical Report CMS-PAS-HIG-14-040, CERN, Geneva, 2015.